Statistics

- Let $\vec{Y} \sim p(\vec{y};q)$. $T(\vec{Y})$ **sufficient statistic** (for the parametric family $p(\vec{y};q)$) ≡ $p(\vec{y};q \,|\, T(\vec{Y})=t)$ is independent of q $\forall t$. ≡ **Neymann-Fisher** factorization: $p(\vec{y};q) = g(t(\vec{y});q)h(\vec{y})$ where g and h are non-negative function.
  - Simple binary hypotheses: $q \in \{q_0, q_1\}$: $t(y) = \dfrac{p(y;q_1)}{p(y;q_0)}$ is a sufficient statistic.

    For $q \in \{q_1, \dots, q_M\}$, $\vec{t}(y) = \left[\dfrac{p(y;q_2)}{p(y;q_1)}, \dots, \dfrac{p(y;q_M)}{p(y;q_1)}\right]$ is a sufficient statistic.
- Sufficient $t(y)$ is a **minimal** sufficient statistic if, for any other sufficient $\tilde{t}$, there is a (measurable) function $h(\cdot)$ such that $t(y) = h(\tilde{t}(y))$.
  - Suppose there exists a function $T(\vec{Y})$ such that for two sample points $\vec{x}$ and $\vec{y}$, the ratio $\dfrac{p(\vec{x};\vec{q})}{p(\vec{y};\vec{q})}$ is a constant as a function of $\vec{q}$ if and only if $T(\vec{x}) = T(\vec{y})$.

    Then $T(\vec{Y})$ is a minimal sufficient statistic for $\vec{q}$.
- A statistic $t(\vec{y})$ is **complete** if $\mathbb{E}\left[g\left(t(\vec{Y})\right)\right] = \int_y g\left(t(\vec{y})\right) p(y;q)dy = 0 \ \forall q \Rightarrow$ $\Pr\left[g\left(t(Y)\right) = 0\right] = 1 \ \forall q \in \Lambda$.
- $T(Y)$ complete and sufficient $\Rightarrow$ minimal sufficient and unique. Ex $X_i \overset{i.i.d}{\sim} \mathcal{U}(0,q)$, $t(\vec{x}) = \max\{x_i\}$.
- ***K*-parameter exponential family**: A family of distributions is said to be a *K*-parameter exponential family if $p(y;q) = \exp\left\{\sum_{i=1}^{K} c_i(q)t_i(y) + d(q) + s(y)\right\} I_A(y)$ $= e^{\sum_{i=1}^{K} c_i(q)t_i(y)} f(q)h(y)I_A(y)$ where $I_A(y)$ is the indicator function not related to q.
  - If $\vec{c}(q) = \left\{\left(c_1(q), \dots, c_K(q)\right), q \in \Lambda\right\}$ has an interior point, then $\vec{t}(y) = \left(t_1(y), \dots, t_K(y)\right)^T$ is complete and sufficient $\Rightarrow$ minimal sufficient and unique

Setup

- Parameter space $\Lambda = \{q\} = \bigcup_{i=0}^{M-1} \Lambda_i$. $p_i = P(\Theta \in \Lambda_i) = \int_{\Lambda_i} \Pr[\Theta = q]dq$. $p_i(q) = \Pr\left[\Theta = q \,|\, \Theta \in \Lambda_i\right]$.
- Hypothesis: $H_i : Y \sim p(y;q)$ $q \in \Lambda_i$, $i = 0, 1, \dots, M-1$.
- Observation space Γ.

- A **deterministic detector** $d: \vec{y} \to \{0, \ldots, M\text{-}1\}$. Partitions the observation space Γ into K disjoint subsets Γ_i and identify Γ_i with Θ_i. When $\Theta_i = \{q_i\}$, $d: \Lambda \to \{q_i\}$.
- A **randomized detector** $\vec{d}(\vec{y}): \vec{y} \to$ pdf/pmf on $\{0,1,\ldots,M-1\}$. $\vec{d}(\vec{y}) = \begin{pmatrix} d_1(\vec{y}) \\ \vdots \\ d_{M-1}(\vec{y}) \end{pmatrix}$.

  $d_k(\vec{y}) = \Pr\left[D = k \middle| \vec{Y} = \vec{y}\right]$. $\sum_{k=0}^{M-1} d_k(\vec{y}) = 1$.
- The detection $D = d$ is a realization according to $D \sim \vec{d}(\vec{y})$.
- Cost: $C(i,q) = \text{Cost } q \to i$. $C(i,j) = \text{Cost } j \to i$. Uniform cost: $C[i,q] = C_{i,j}, q \in \Lambda_j$.

  Assume $C_{ij} > C_{ii}$ for all i, j.

The Bayesian Detector

- Bayesian / Bayes Risk:

  $$R(d) = E[\text{Cost}] = \int_q p(q) R_q(d) = \sum_{i=0}^{M-1} p_i R_i(d) = \int p(\vec{y}) R(d|\vec{y}) d\vec{y}.$$
- Conditional risk: $R(d|\vec{y}) = E\left[\text{Cost} \middle| \vec{Y} = \vec{y}\right]$
- $$R_q(d) = \sum_{i=0}^{M-1} C(i,q) \Pr\left[D = i \middle| \Theta = q\right] = \sum_{i=0}^{M-1} C(i,q) \int d_i(y) p(y;q) dy$$

  $$= \sum_{i=0}^{M-1} C(i,q) \Pr\left[y \in \Gamma_i \middle| \Theta = q\right] = \sum_{i=0}^{M-1} C(i,q) \int_{\Gamma_i} p(y;q) dy \text{ for deterministic detector}$$
- $$R_k(d) = \int_{q \in \Lambda_k} p_k(q) R_q(d) dq.$$
- Bayesian Detector $d_B = \underset{d}{\operatorname{argmin}}\, R(d)$. Because the distribution of y doesn't depends on the detector, the Bayesian detector for a given $\vec{y}$:

  $$\boxed{d_B(y) = \underset{d}{\operatorname{argmin}}\, R(d|\vec{y}) = \underset{d}{\operatorname{argmin}}\, E\left[\text{Cost} \middle| \vec{Y} = \vec{y}\right]}.$$
- **Simple Hypotheses**: $q \in \{q_1, \ldots, q_{M-1}\}$
  - The Bayesian detector is deterministic: $d_k(\vec{y}) = \Pr\left[D = k \middle| \vec{Y} = \vec{y}\right] = \begin{cases} 1, & k = k_0 \\ 0, & o/w. \end{cases}$,

    where $\boxed{k_0 = d(y) = \underset{k}{\operatorname{argmin}}\, E\left[\text{Cost} \middle| D = k, \vec{Y} = \vec{y}\right] = \underset{k}{\operatorname{argmin}} \sum_{j=0}^{M-1} C_{kj} p_j p(\vec{y}; q_j)}$.
    - $$E\left[\text{Cost} \middle| D = k, \vec{Y} = \vec{y}\right] = \sum_{j=0}^{M-1} C_{kj} p\left(\Theta = q_j \middle| \vec{Y} = \vec{y}\right) = \frac{1}{p(\vec{y})} \sum_{j=0}^{M-1} C_{kj} p_j p(\vec{y}; q_j).$$
  - $$R_q(d) = \sum_{i=0}^{M-1} C(i,q) \Pr\left[D = i \middle| \Theta = q\right] = \sum_{i=0}^{M-1} C(i,q) \int_{\Gamma_i} p(y;q) dy.$$

- $\boxed{R(d)=\sum_{j=0}^{M-1} p_j \sum_{k=0}^{M-1} C_{kj} \int_{\Gamma_k} p(y;q_j)\,dy}$. With uniform cost, it is the probability of error.

- **Binary Simple Hypotheses**: $p_0 = \Pr[q = q_0]$, $p_1 = 1 - p_0$.
  - d_{B,p_0} = Bayesian detector for prior p_0
  - $d(y) = d(y)$. $d(\bar{y}) = \begin{cases} 1, & \dfrac{p(\bar{y}|q_1)}{p(\bar{y}|q_0)} \geq t \\ 0, & \text{otherwise} \end{cases}$ where $t = \dfrac{(C_{10}-C_{00})p_0}{(C_{01}-C_{11})p_1}$.
    - $\left(C_{00}p(\bar{y}|q_0)p_0 + C_{01}p(\bar{y}|q_1)p_1 \geq C_{10}p(\bar{y}|q_0)p_0 + C_{11}p(\bar{y}|q_1)p_1\right)$
  - $\Gamma_1 = \{y : p(\bar{y}|q_1) \geq t\, p(\bar{y}|q_0)\} = \Gamma_0^c$.
  - $R(d) = C_{10}p_0 \int_{\Gamma_1} p(y;q_0)\,dy + C_{01}p_1 \int_{\Gamma_0} p(y;q_1)\,dy + C_{00}p_0 \int_{\Gamma_0} p(y;q_0)\,dy + C_{11}p_1 \int_{\Gamma_1} p(y;q_1)\,dy$.
  - Ex. $H_0 : Y \sim \mathcal{N}(q_0, s^2)$, $H_1 : Y \sim \mathcal{N}(q_1, s^2)$. $q_0 < q_1$.

    Then, $\Gamma_1 = \left\{ y : y \geq g = \dfrac{q_0+q_1}{2} + \dfrac{s^2}{q_1-q_0} \ln \dfrac{(C_{10}-C_{00})p_0}{(C_{01}-C_{11})p_1} \right\}$. For uniform cost,

    $R(d) = p_0 Q\left(\dfrac{g-q_0}{s}\right) + p_1 Q\left(\dfrac{q_1-g}{s}\right)$.
  - $0 \leq a \leq t = \dfrac{p_0}{1-p_0} \leq b \Leftrightarrow \dfrac{a}{1+a} \leq p_0 \leq \dfrac{b}{1+b}$

- **Uniform cost with identity cost matrix Composite Binary Hypothesis Testing**
  - $p(\bar{y}|\Theta \in \Lambda_i) = \int_{\Lambda_i} p(\bar{y}|q)\Pr[\Theta = q|\Theta \in \Lambda_i]dq = \int_{\Lambda_i} p(\bar{y}|q)\dfrac{\Pr[\Theta = q]}{p_i}dq$

    $= \dfrac{1}{p_i} \int_{\Lambda_i} p(\bar{y}|q)\Pr[\Theta = q]dq$
    - $p_i(q) = \Pr[\Theta = q|\Theta \in \Lambda_i] = \begin{cases} \dfrac{\Pr[\Theta = q]}{p_i} & q \in \Lambda_i \\ 0 & q \notin \Lambda_i \end{cases}$
  - $d_B(y) = \begin{cases} 1, & L(y) \geq t \\ 0, & L(y) < t \end{cases}$. $t = \dfrac{(C_{10}-C_{00})p_0}{(C_{01}-C_{11})p_1}$ where $p_i = P(\Theta \in \Lambda_i) = \int_{\Lambda_i} \Pr[\Theta = q]\,dq$.

- $L(y) = \dfrac{p(\bar{y}|\Theta\in\Lambda_1)}{p(\bar{y}|\Theta\in\Lambda_0)} = \dfrac{\frac{1}{p_1}\int_{\Lambda_1} p(\bar{y}|q)\Pr[\Theta=q]dq}{\frac{1}{p_0}\int_{\Lambda_0} p(\bar{y}|q)\Pr[\Theta=q]dq} \geq t = \dfrac{(C_{10}-C_{00})p_0}{(C_{01}-C_{11})p_1}$.
- $R(d) = \Pr[D=1|\Theta\in\Lambda_0]p_0 + \Pr[D=0|\Theta\in\Lambda_1]p_1$

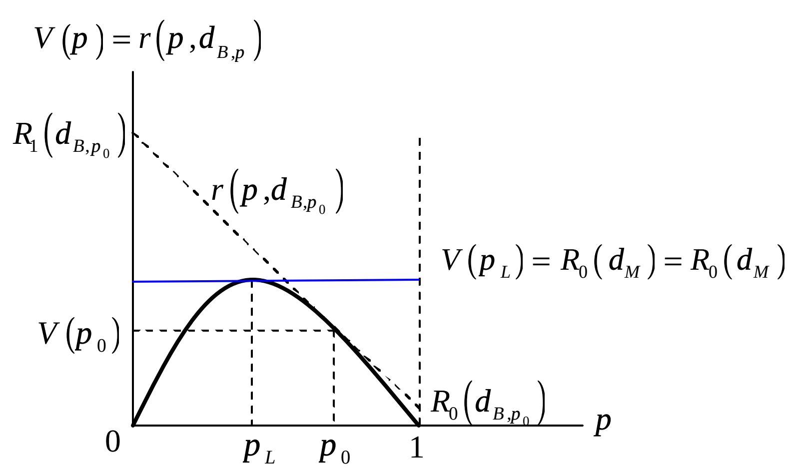
$$= p_0 \int_{q\in\Lambda_0}\int_{\Gamma_1} p(y|q)dy\, p(q|q\in\Lambda_0)dq + p_1 \int_{q\in\Lambda_1}\int_{\Gamma_0} p(y|q)dy\, p(q|q\in\Lambda_1)dq$$

$$= \int_{\Gamma_1}\int_{q\in\Lambda_0} p(y|q)p(q)dq\,dy + \int_{\Gamma_0}\int_{q\in\Lambda_1} p(y|q)p(q)dq\,dy$$

$$= \int_{q\in\Lambda_0}\int_{\Gamma_1} p(y|q)p(q)dq\,dy + \int_{q\in\Lambda_1}\int_{\Gamma_0} p(y|q)p(q)dq\,dy$$

Minimax Detection

- Minimax detector/rule/criterion: $\min_d \max_q R_q(d)$.
- If we know $\Theta \sim p_i(q)\ \forall q\in\Lambda_i\ \forall i$ then minimax detector is $\min_d \max_k R_k(d)$.
  - $R_k(d) = \int_{q\in\Lambda_k} p_k(q) R_q(d) dq$.
- **Simple binary hypothesis testing**:
  - Risk for d given prior $p_0 = p$: $\boxed{r(p,d) = p R_{q_0}(d) + (1-p) R_{q_1}(d)}$. (Linear wrt. p)
    - $\forall p\ \forall d\ r(p,d) \geq r(p, d_{B,p})$.
  - Minimax detector is $d_M = \min_d \max\{R_0(d), R_1(d)\}$.
  - Minimum Bayesian risk $V(p_0) = r(p_0, d_{B,p_0})$



  - Concave and continuous in [0,1].

- $\exists p_L = \arg\max_p V(p)$ = **least favorable prior**.
- If $V'(p)$ exists at $p = p^*$, then $r(p, d_{B,p^*})$ is a tangent line of $V(p)$ at p^*.

  $V'(p^*) = R_0(d_{B,p^*}) - R_1(d_{B,p^*})$.
- $R_0(d) = E[\text{Cost}|q_0] = \int_y C(1,q_0) d(y) p(y;q_0) dy + \int_y C(0,q_0)(1-d(y)) p(y;q_0) dy$

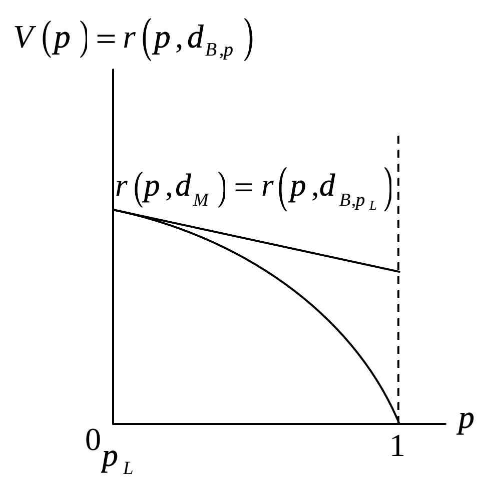
  $R_1(d) = E[\text{Cost}|q_1] = \int_y C(1,q_1) d(y) p(y;q_1) dy + \int_y C(0,q_1)(1-d(y)) p(y;q_1) dy$
- $d_{B,p}$ = the Bayesian detector designed at p.

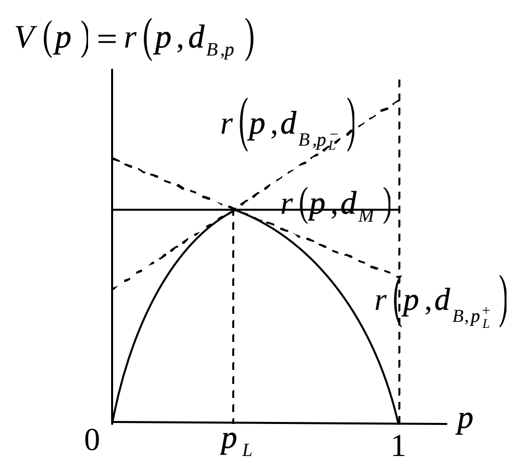
  $d_{B,p} = \arg\min_d r(p,d)$.

  Minimum Bayesian risk given prior: $n(p) = r(p, d_{B,p}) \le r(p,d), \forall d$.

- Solving for Minimax detector for binary simple hypotheses:
  - Equalizer rule: If there exists a prior p_L such that $\boxed{R_{q_0}(d_{B,p_L}) = R_{q_1}(d_{B,p_L})}$, then p_L is the least favorable prior $(\forall p\ V(p_L) \ge V(p))$, and the minimax detector $d_M = d_{B,p_L}$.
  - If not then,
    - If $p_L = 0$ or 1, then $d_M = d_{B,p_L}$.
    - Otherwise, (if $V(p)$ is linear in the small neighborhood of p_L,) consider d_{B,p_L^-} and d_{B,p_L^+}, $d_M(y) = \begin{cases} d_{B,p_L^-}(y) & \text{with probability } q \\ d_{B,p_L^+}(y) & \text{with probability } 1-q \end{cases}$ with q from $R_0(d_M) = R_1(d_M)$.

$$q = \frac{R_0(d_{B,p_L^+}) - R_1(d_{B,p_L^+})}{\left(R_0(d_{B,p_L^+}) - R_1(d_{B,p_L^+})\right) - \left(R_0(d_{B,p_L^-}) - R_1(d_{B,p_L^-})\right)} = \frac{V'(p_L^+)}{V'(p_L^+) - V'(p_L^-)}.$$

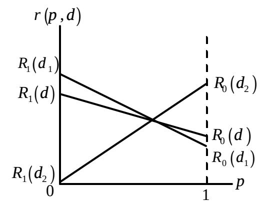




- The minimax risk = $V(p_L) = R_{q_0}(d_{B,p_L}) = R_{q_1}(d_{B,p_L})$ for all cases.
- A detector d_M is minimax if it satisfies the equalizer rule $R_{q_0}(d_M) = R_{q_1}(d_M)$ and there exists a prior p_L such that $V(p_L) = R_{q_0}(d_M)$.

- Randomized Bayesian Detector:

$$d(y) = \begin{cases} d_1(y) & \text{with probability } q \\ d_2(y) & \text{with probability } 1-q \end{cases}.$$



- $R_i(d) = qR_i(d_1) + (1-q)R_i(d_2)$

- The equalizer rule remains valid for composite hypotheses. If the min Bayesian risk is differentiable at the least favorable prior, then the Bayesian detector is the minimax detector.

Neyman-Pearson Detector

- $\min_d R_{q_K}(d)$ subject to $R_{q_i}(d) \le a_i$, $i < K$.
- Binary Hypotheses: $H_i : Y \sim p(y;q)$, $q \in \Lambda_i$, $i = 0,1$. H_0: null hypothesis. H_1: the alternative.
- Simple binary hypotheses: $H_0 : Y \sim p(y;q_0)$; $H_1 : Y \sim p(y;q_1)$. Γ_0: acceptance region; Γ_1: rejection region.
- **False alarm** (I) $P_F(d;q) = \Pr[D=1;q] = \int d(y)p(y;q)dy = E_q(d(y))$ $\forall q \in \Lambda_0$.
- The **size** / level of a detector $a(d) = \sup_{q \in \Lambda_0} P_F(d;q)$.
- **Miss detection** (II) $P_M(d;q) = \Pr[D=0;q] = \int(1-d(y))p(y;q)dy = 1 - E_q(d(y))$ $\forall q \in \Lambda_1$.
- The **power** of a detector = $P_D(d;q) = \Pr[D=1;q] = \int d(y)p(y;q)dy = E_q(d(y))$ $\forall q \in \Lambda_1$.
- $P_D(d;q) = 1 - P_M(d;q)$ $\forall q \in \Lambda_1$. | $P_F(d;q)$ and $P_D(d;q)$ has the same formula but using q from different sets.
- **Uniformly most powerful** (UMP): A size a detector d_{UMP} is UMP if $\forall d$ of size $\le a$, $P_D(d_{UMP};q) \ge P_D(d;q)$ $\forall q \in \Lambda_1$. ($\forall d$ $a(d) \le a(d_{UMP}) \Rightarrow P_D(d_{UMP};q) \ge P_D(d;q)$ $\forall q \in \Lambda_1$.)
  - For simple binary hypotheses, $d_{UMP} = \arg\max_{d,\; P_M(d;q_0)\le a} P_D(d;q_1)$.
  - (If $E_{q^*}[d_{UMP}(y)] = \Pr[D=1;q^*] = a$, then $d_{UMP}(y)$ is a size a NP detector for $H_0 : Y \sim p(y;q^*)$; $H_1 : Y \sim p(y;q_1)$ for any $q_1 \in \Lambda_1$.?)
- For simple binary hypotheses, **NP detector** is $\arg\max_{d,\; P_F(d)\le a} P_D(d) = \arg\min_{d,\; P_F(d)\le a} P_M(d) = \arg\min_{d,\; R_0(d)\le a} R_1(d)$ with uniform cost.

- **Neyman-Pearson Lemma** for simple binary hypotheses:
  1) Optimality. Any $d^*(y) = \begin{cases} 1, & p(y;q_1) > h\, p(y;q_0) \\ g(y), & p(y;q_1) = h\, p(y;q_0) \\ 0, & p(y;q_1) < h\, p(y;q_0) \end{cases}$ for some $h \geq 0$ and $g(y) \in [0,1]$ is the best of its size.

     $P_D(d^*) - P_D(d) > h\left(P_F(d^*) - P_F(d)\right)$ for all d.
  2) Existence. $\forall a \in [0,1]$, there exists a detector of the form above.

     $h = \min_{\Pr[L(y) > h_0; q_0] \leq a} h_0$. $g(y) = \begin{cases} g_0, & \Pr[L(y) = h; q_0] \neq 0 \\ \text{arbitrary}, & \text{otherwise} \end{cases}$.

     $\Pr[L(y) > h_0; q_0]$ is a complimentary distribution function, right continuous, and monotonically decreasing.

     $$\boxed{g_0 = \frac{a - \Pr[L(y) > h; q_0]}{\Pr[L(y) = h; q_0]}}.$$

     ($a = P_F = \Pr[L(y) > h; q_0] + g \Pr[L(y) = h; q_0]$.)
  3) Uniqueness. If d' is a size a NP detector, then $d'(y)$ has the form above except perhaps for a set of y with zero probability under both H_0 and H_1.
  - Note: 1) $P_D(d^*) = \Pr[D = 1; q_1] = \Pr[L(y) > h; q_1] + g \Pr[L(y) = h; q_1]$. Since $\Pr[L(y) > h; q_1]$ is also monotonically decreasing, we want low h to get high $P_D(d^*)$. 2) Helpful to plot $\Pr[L(y) > h; q_0]$ vs. h. 3) Can work with $t(y)$ in stead of $L(y)$ when the transformation is 1:1, increasing.

UMP detector

- Let q be a real parameter. The real-parameter family $p(y;q)$ has **monotone likelihood ratio** (in $T(y)$) if $\forall q < q'$, $p(y;q)$ and $p(y;q')$ are distinct and $L(y;q',q) = \frac{p(y;q')}{p(y;q)}$ is a nondecreasing function of some real valued $T(y)$.
  - Ex. 1) $Ce^{-\frac{\|\bar{y} - q\bar{m}\|}{2s^2}}$. $T(\bar{y}) = \bar{y}^T \bar{u}$. 2) $c(q) h(\bar{y}) e^{Q(q) T(\bar{y})}$ when $Q(q)$ is monotone. 3) i.i.d. Bernoulli, $T(\bar{y}) = \sum_k y_k$.
- One-sided Hypotheses Testing: $H_0 : Y \sim p(y;q)\ q \leq q_*$. $H_1 : Y \sim p(y;q)\ q > q_*$ (or $q > q_1 \geq q_*$).
  - **The Kalin Rubin Theorem**: Let q be a real parameter and let $p(y;q)$ have monotone likelihood ration in $T(y)$. For testing the one-sided hypotheses, there

exists a size a UMP detector of the form $d^*(y) = \begin{cases} 1, & T(y) > t \\ g, & T(y) = t \\ 0, & T(y) < t \end{cases}$ where t and g are determined by the size constraint $t = \min_{\Pr[T(y)>t_0;q_*]\le a} t_0$, $E_{q_*}[d(y)] = \int d(y)\,p(y;q_*)\,dy = a$. t and g are functions of q_*.

- Given any $q_1 < q_2$, and d^* with NP form. Then $P_F(d^*;q_2) \ge P_F(d^*;q_1)$. $\mathbb{E}_q[d^*(y)] = \Pr[D=1|q]$ is a nondecreasing $(\uparrow)$ function of q. So, for $\Lambda_0 = (-\infty, q^*]$, the size (false alarm) of d^* is $a(d^*) = \sup_{q\in\Lambda_0} P_F(d^*;q) = P_F(d^*;q^*)$.
- Two-sided Hypothesis: $p(y;q) = h(y)e^{a(q)T(y)-b(q)}$ with nondecreasing $(\uparrow)$ $a(q)$. $\exists$ UMP detector for $H_0: q \le q_1$ or $q \ge q_2$, $H_1: q \in (q_1, q_2)$ of the form $d^*(y) = \begin{cases} 1, & c_1 < T(y) < c_2 \\ g_i, & T(y) = c_i \\ 0, & \text{otherwise} \end{cases}$ where $c_1 < c_2$ and g_i are determined by $\mathbb{E}_{q_1}[d^*(y)] = \mathbb{E}_{q_2}[d^*(y)] = a$.
- If H_0 is surrounded by H_1, then suspect no UMP detector.

Bayesian Estimaion

- Estimate random $\Theta \sim p(q)$ from $Y \sim p(y|q)$. The cost is $E\left[C\left(\hat{\Theta} - \Theta\right)\right]$. Bayesian estimator $\hat{q}$ minimize $E\left[C\left(\hat{q}(Y) - \Theta\right)\right]$. $\hat{q}_{Bayesian}(y) = \operatorname{argmin}_{\hat{q}} R(\hat{q}|y)$. $R(\hat{q}|y) = E\left[C\left(\hat{q} - \Theta\right)\middle|Y = y\right] = \int C(\hat{q} - q)\,p(q|y)\,dq$.
- **MMSE**: $\hat{q}_{MMSE}(y) = \operatorname{argmin}_{\hat{q}} \int \left\|\hat{q} - q\right\|^2 p(q|y)\,dq = E[\Theta|Y = y]$.
  - For jointly **Gaussian** $\vec{\Theta}$ and $\vec{Y}$, $\hat{q}_{MMSE,\mathcal{N}}(y) = \vec{m}_{\vec{\Theta}} + \Lambda_{\vec{\Theta}\vec{Y}}(\Lambda_{\vec{Y}\vec{Y}})^{-1}(\vec{y} - \vec{m}_{\vec{Y}})$. In addition, if $\vec{m}_{\vec{\Theta}}, \vec{m}_{\vec{Y}}$ are zero, then $\hat{q}_{MMSE,\mathcal{N},0}(y) = \Lambda_{\vec{\Theta}\vec{Y}}(\Lambda_{\vec{Y}\vec{Y}})^{-1}\vec{y}$, linear. For example, let $Y_k = a_k\Theta + N_k$, $k = 1, \ldots, n$. $\Theta \sim \mathcal{N}(0, s_\Theta^2) \perp N_k \overset{i.i.d.}{\sim} \mathcal{N}(0, s_N^2)$. Then $\hat{q}_{MMSE,\mathcal{N},0}(y) = \dfrac{1}{\left(\vec{a}^T\vec{a} + \dfrac{n}{g}\right)}\vec{a}^T\vec{y}$ where $g = \dfrac{ns_\Theta^2}{s_N^2}$. If $\vec{a} = \vec{1}$, then $\hat{q}_{MMSE,\mathcal{N},0}(y) = \dfrac{g}{1+g}\dfrac{1}{n}\sum_i y_i$.
  - <u>Linear observation model</u>: Let $\vec{Y} = H\vec{S} + \vec{W}$, $\vec{S} \perp \vec{W}$, $\vec{S} \sim \mathcal{N}(\vec{m}_{\vec{S}}, \Sigma_{\vec{S}})$, $\vec{W} \sim \mathcal{N}(0, \Sigma_{\vec{W}})$, then the MMSE estimator is

$\mathbb{E}\left[\vec{S}\,\middle|\,\vec{Y}\right] = \vec{m}_{\vec{S}} + \Sigma_{\vec{S}} H^H \left(H\Sigma_{\vec{S}} H^H + \Sigma_{\vec{W}}\right)^{-1} \left(\vec{Y} - H\vec{m}_{\vec{S}}\right)$. Error covariance matrix: $Cov\left[\vec{S}\,\middle|\,\vec{Y}\right] = \Sigma_{\vec{S}} - \Sigma_{\vec{S}} H^H \left(H\Sigma_{\vec{S}} H^H + \Sigma_{\vec{W}}\right)^{-1} H\Sigma_{\vec{S}}$.

- Given y, assume 1) $C\left(\hat{q} - q\right)$ is symmetrical, i.e. $C(x) = C(-x)$, and convex ∪, 2) let $\hat{q}_m = E\left(\Theta\middle|y\right)$, then $p\left(q\,|y\right)$ is symmetrical with respect to $\hat{q}_m$, i.e. $p\left(\hat{q}_m + q\,|y\right) = p\left(\hat{q}_m - q\,|y\right)$, then $\hat{q}_{Bayesian}(y) = \hat{q}_{MMSE}(y) = \hat{q}_m$.

- $\hat{q}_{MMAE}(y) = \underset{\hat{q}}{\operatorname{argmin}} \int \left\|\hat{q} - q\right\|_1 p\left(q|y\right) dq = \underset{\hat{q}}{\operatorname{argmin}} \sum_{i=1}^{n} \int \left|\hat{q}_i - q_i\right| p\left(q_i\,|y\right) dq_i$ $= \mathbb{M}\left[\Theta\middle|Y = y\right] \left(\int_{-\infty}^{\hat{q}_i} p\left(q_i\,|y\right) dq_i = \int_{\hat{q}_i}^{\infty} p\left(q_i\,|y\right) dq_i \;\; \forall i\right)$.

- For $C_{u,\Delta}\left(\hat{q} - q\right) = \begin{cases} 1, & \left\|\hat{q} - q\right\|_\infty > \dfrac{\Delta}{2} \\ 0, & \left\|\hat{q} - q\right\|_\infty \le \dfrac{\Delta}{2} \end{cases}$, $\hat{q}_{u,\Delta}(y) = \underset{\hat{q}}{\operatorname{argmin}}\, 1 - \int_{\hat{q}_1 - \frac{\Delta}{2}}^{\hat{q}_1 + \frac{\Delta}{2}} \cdots \int_{\hat{q}_n - \frac{\Delta}{2}}^{\hat{q}_n + \frac{\Delta}{2}} p\left(q\,|y\right) dq$ = $\underset{\hat{q}}{\operatorname{argmax}} \int_{\hat{q}_1 - \frac{\Delta}{2}}^{\hat{q}_1 + \frac{\Delta}{2}} \cdots \int_{\hat{q}_n - \frac{\Delta}{2}}^{\hat{q}_n + \frac{\Delta}{2}} p\left(q\,|y\right) dq$. Let $\Delta \to 0$, then $\hat{q}_{MAP}(y) = \underset{\hat{q}}{\operatorname{argmax}}\, p\left(q\,|y\right)$.

- **MSE** For $\mathbb{E}\left[\vec{X} - \hat{X}\right] = 0$ unbiased, $Cov\left(\vec{X} - \hat{X}\right) = \mathbb{E}\left[\left(\vec{X} - \hat{X}\right)\left(\vec{X} - \hat{X}\right)^H\right]$. $MSE\left(\hat{X}\right) = \mathbb{E}\left[\left\|\vec{X} - \hat{X}\right\|^2\right] = \sum_{k=1}^{n} \mathbb{E}\left[\left|X_i - \hat{X}_i\right|^2\right] = \operatorname{trace}\left(Cov\left(\vec{X} - \hat{X}\right)\right)$.

- **Linear MMSE**: $\mathbb{E}\vec{\Theta} = 0$?. Given zero mean random variables Y_i, $i = 1, \ldots, n$,
  - $\hat{\Theta} = \vec{f}^H \vec{Y} = \sum_{i=1}^{n} f_i^* Y_i$. $\vec{f}^H = \mathbb{E}\left[\Theta\vec{Y}^H\right]\left(\mathbb{E}\left[\vec{Y}\vec{Y}^H\right]\right)^{-1}$ minimizes **MSE** = $\mathbb{E}\left[\left|\Theta - \hat{\Theta}\right|^2\right]$ to $\mathbb{E}\left[\left|\Theta\right|^2\right] - \mathbb{E}\left[\Theta\vec{Y}^H\right]\left(\mathbb{E}\left[\vec{Y}\vec{Y}^H\right]\right)^{-1}\mathbb{E}\left[\vec{Y}\Theta^H\right]$.
    - $\hat{\Theta}$ is the orthogonal projection of Θ onto $\operatorname{span}\{Y_1, \ldots, Y_n\}$.
    - $\mathbb{E}\left[\hat{\Theta} - \vec{\Theta}\right] = 0$
  - For $\vec{Y} \in \mathcal{C}^n$, $\vec{X} \in \mathcal{C}^m$, $F^H = \mathbb{E}\left[\vec{X}\vec{Y}^H\right]\left(\mathbb{E}\left[\vec{Y}\vec{Y}^H\right]\right)^{-1}$.
    - $T\hat{X}$ is the linear MMSE estimate of $T\vec{X}$.
    - $\hat{X}$ is also the optimal linear estimate using the weighted cost function $\min_{F^H} \mathbb{E}\left[\left(\vec{X} - F^H\vec{Y}\right)^H \Lambda \left(\vec{X} - F^H\vec{Y}\right)\right]$ for any $\Lambda \ge 0$.

- Linear observation model: $\vec{Y} = H\vec{S} + \vec{W}$, $\mathbb{E}\left[\begin{pmatrix}\vec{S}\\ \vec{W}\end{pmatrix}\right] = 0$,

  $Cov\left(\begin{pmatrix}\vec{S}\\ \vec{W}\end{pmatrix}, \begin{pmatrix}\vec{S}\\ \vec{W}\end{pmatrix}\right) = \begin{bmatrix}\Sigma_{\vec{S}\vec{S}} & 0\\ 0 & \Sigma_{\vec{W}\vec{W}}\end{bmatrix}$, then the linear MMSE estimate of $\vec{S}$ is given by

  $\hat{S} = \Sigma_{\vec{S}\vec{S}} H^H \left(H\Sigma_{\vec{S}\vec{S}} H^H + \Sigma_{\vec{W}\vec{W}}\right)^{-1} \vec{Y}$.

- **Affine MMSE estimator**: Given random vector $\vec{Y}$. If $\mathbb{E}\left[\vec{Y}\right]$ and $\mathbb{E}\left[\vec{X}\right]$ are known, then the MMSE affine estimator of $\vec{X}$ by $\vec{Y}$ is given by $\hat{X} = m_{\vec{X}} + \Sigma_{\vec{X}\vec{Y}}\Sigma_{\vec{Y}}^{-1}\left(\vec{Y} - m_{\vec{Y}}\right)$.

  $Cov\left(\vec{X} - \hat{X}\right) = \Sigma_{\vec{X}} - \Sigma_{\vec{X}\vec{Y}}\Sigma_{\vec{Y}}^{-1}\Sigma_{\vec{Y}\vec{X}}$. $\mathbb{E}\left[\vec{X} - \hat{X}\right] = 0$.

  $Cov\left(\hat{X}, \hat{X}\right) = Cov\left(\vec{X}, \hat{X}\right) = \Sigma_{\vec{X}\vec{Y}}\Sigma_{\vec{Y}}^{-1}\Sigma_{\vec{Y}\vec{X}}$.

  - If T is nonsingular and $\vec{u} = T\vec{y}$, then the MMSE affine estimator of $\vec{x}$ using $\vec{y}$ is the same as that using $\vec{u}$.
  - If $\vec{W}$ and $\begin{pmatrix}\vec{X}\\ \vec{Y}\end{pmatrix}$ are uncorrelated, and $\vec{V} = B\vec{X} + \vec{W}$, then the MMSE affine estimator of $\vec{V}$ using $\vec{Y}$ is given by $\underset{\text{affine,MMSE}}{\hat{V}\left(\vec{Y}\right)} = B\hat{X} = Bm_{\vec{X}} + B\Sigma_{\vec{X}\vec{Y}}\Sigma_{\vec{Y}}^{-1}\left(\vec{Y} - m_{\vec{Y}}\right)$,

    $Cov\left(\vec{V} - \hat{V}\right) = B\left(Cov\left(\vec{X} - \hat{X}\right)\right)B^H + \Sigma_{\vec{W}}$

Point Estimation

- Criterion: Minimize **MSE** (risk) $\underset{\text{for a fixed } q}{M_q\left(\hat{g}\right)} = \mathbb{E}\left[\left\|\hat{g}\left(\vec{Y}\right) - g\left(q\right)\right\|^2\right]$.
- $\mathbb{E}\left[\left\|\hat{g}\left(\vec{Y}\right) - g\left(q\right)\right\|^2\right] = \mathbb{E}\left[\left\|\hat{g}\left(\vec{Y}\right) - \mathbb{E}\left[\hat{g}\left(\vec{Y}\right)\right]\right\|^2\right] + \left\|\mathbb{E}\left[\hat{g}\left(\vec{Y}\right)\right] - g\left(q\right)\right\|^2$
- An estimator $\hat{g}\left(\vec{y}\right)$ of $g\left(q\right)$ is **unbiased** if $E\left[\hat{g}\left(\vec{Y}\right)\right] - g\left(q\right) = 0\ \forall q$. Then,

  $M_q\left(\hat{g}\right) = \mathbb{E}\left[\left\|\hat{g}\left(\vec{Y}\right) - \mathbb{E}\left[\hat{g}\left(\vec{Y}\right)\right]\right\|^2\right]$. ; For unbiased $\hat{q}$, $M\left(\hat{q}\right) = \text{trace}\left\{Cov\left(\hat{q}\left(Y\right)\right)\right\}$.

  Note that $Cov\left(\hat{q}\left(Y\right)\right) = Cov\left(\hat{q}\left(Y\right) - q\right) = Cov\left(\hat{q}\left(Y\right) + \vec{a}\right)$ always.

UMVU

- An estimator $\hat{g}\left(\vec{y}\right)$ of $g\left(q\right)$ is **UMVU** (uniformly minimum variance unbiased) if 1) unbiased. 2) For all unbiased $\hat{g}'$, $\forall q$ $M_q\left(\hat{g}\right) \le M_q\left(\hat{g}'\right)$.
- **Rao-Blackwell Theorem**: Suppose that $T\left(\vec{Y}\right)$ is sufficient for q and that $\hat{g}\left(\vec{y}\right)$ is an estimator for $g\left(q\right)$ with $E\left[\left|\hat{g}\left(\vec{y}\right)\right|_1\right] < \infty$ for all q.

Let $\hat{g}^*(\vec{y}) = \mathbb{E}\left[\hat{g}(\vec{y}) \middle| T(\vec{Y}) = T(\vec{y})\right]$. Then 1)

$\hat{g}^*(\vec{y}) = \int\limits_{\substack{\vec{y}' \\ T(\vec{y}')=T(\vec{y})}} \hat{g}(\vec{y}')\, p_{\vec{Y}|T(\vec{Y})}\left(\vec{y}' \middle| T(\vec{y})\right) dy'$ 2)

$\forall q\, , \mathbb{E}\left[\left\|\hat{g}^*(\vec{Y}) - g(q)\right\|^2\right] \le \mathbb{E}\left[\left\|\hat{g}(\vec{Y}) - g(q)\right\|^2\right]$. 3) If components of $\hat{g}$ have finite variances, then the strict inequality holds unless $\Pr\left[\hat{g}^*(\vec{Y}) = \hat{g}(\vec{Y})\right] = 1$.

- $\mathbb{E}\left[\left\|\hat{g}(\vec{Y}) - g(q)\right\|^2\right] = \mathbb{E}\left[\left\|\hat{g}(\vec{Y}) - \hat{g}^*(\vec{Y})\right\|^2\right] + \mathbb{E}\left[\left\|\hat{g}^*(\vec{Y}) - g(q)\right\|^2\right]$

Furthermore, if $\hat{g}(\vec{y})$ is unbiased, then 4) $\hat{g}^*(\vec{y})$ is unbiased for $g(q)$. 5) (2) can be written as $\forall q \quad Var\left[\hat{g}^*(\vec{Y})\right] \le Var\left[\hat{g}(\vec{Y})\right]$.

- If $\hat{g}(\vec{y}) = h\left(T(\vec{y})\right)$, then $\hat{g}^*(\vec{y}) = \hat{g}(\vec{y})$

- **Lehmann-Scheffé Theorem**: If $T(Y)$ is complete sufficient, and $\hat{g}(Y)$ is any unbiased estimator of $g(q)$. Then $\boxed{\hat{g}^*(T(y)) = \mathbb{E}\left[\hat{g}(Y) \middle| T(Y) = T(y)\right]}$ is an **UMVU** estimator.
- Shortcut: Knowing $T(Y)$ is complete sufficient, try finding $\mathbb{E}\left[T(Y)\right]$.
- For one-parameter exponential family, $T(Y)$ is complete and sufficient, if it is unbiased, then it is UMVU.

CRB: Cramér-Rao Lower Bound

- The <u>**score function**</u> $s(y;q) = \begin{bmatrix} \frac{\partial}{\partial q_1} \ln p(y;q) \\ \vdots \\ \frac{\partial}{\partial q_K} \ln p(y;q) \end{bmatrix} = \nabla_q \ln p(y;q)$. $\mathbb{E}_q\left[s(Y;q)\right] = 0$.
- **Fisher Information Matrix**: $I(q) = E\left[s(Y;q)s'(Y;q)\right] = Cov\left[s(Y;q)\right] \ge 0$.

$I_{ij}(q) = \mathbb{E}\left[\frac{\partial}{\partial q_i} \ln p(Y;q) \frac{\partial}{\partial q_j} \ln p(Y;q)\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial q_i \partial q_j} \ln p(Y;q)\right]$

$I(q) = \mathbb{E}\left[\left(\nabla_q \ln p(Y;q)\right)\left(\nabla_q \ln p(Y;q)\right)^T\right] = -\mathbb{E}\left[\nabla_q^2 \ln p(Y;q)\right]$.

For scalar q, $I(q) = E\left[\frac{\partial}{\partial q} \ln p(y;q) \frac{\partial}{\partial q} \ln p(y;q)\right] = -E\left[\frac{\partial^2}{\partial q^2} \ln p(y;q)\right]$

- The **Cramér-Rao Bound**: Let $\hat{q}$ be a scalar <u>unbiased</u> estimator of q. Then, <u>CRLB</u>: $M(\hat{q}) = Var(\hat{q}) = Var\left(\hat{q}(Y) - q\right) = \mathbb{E}\left[\left(\hat{q}(Y) - q\right)^2\right] \ge \frac{1}{I(q)}$ with equality iff

$s(y;q) = \frac{\partial}{\partial q} \ln p(y;q) = I(q)\left(\hat{q} - q\right)$.

- **Information lower bound**: For biased estimator, $\mathbb{E}\left[\hat{q}(Y)\right]=\Phi(q)$, then $Var\left(\hat{q}(Y)\right)\geq\frac{\left(\Phi'(q)\right)^2}{I(q)}$ with equality iff $s(x;q)=I(q)\left(\hat{q}-\Phi(q)\right)$.
- If $\hat{q}(Y)$ achieves information lower bound, then it has minimum variance among all estimators $\tilde{q}(Y)$ satisfying $\frac{\partial}{\partial q}\mathbb{E}\left[\tilde{q}(Y)\right]=\frac{\partial}{\partial q}\mathbb{E}\left[\hat{q}(Y)\right]$. Furthermore, if $\hat{q}(Y)$ is unbiased, then $\hat{q}(Y)$ is efficient and UMVU.
- <u>One parameter exponential family</u>: Let Λ be an open interval, and $p(y;q)$ = $C(q)e^{g(q)T(y)}h(y)$. Within regularity, 1) the information lower bound is achieved by $\hat{q}(Y)$ if any only if $\hat{q}(Y)=T(Y)$. Also, 2) $T(Y)$ is complete and sufficient. 3) If $T(Y)$ is unbiased, then it is <u>UMVU and efficient</u>.
- An unbiased estimator is **efficient** if it achieves CRB.
  - An efficient estimator is UMVU but an UMVU estimator may not be efficient (when CRB is not achievable.)
- If $\hat{q}(Y)$ achieves CRLB, then it is the solution to the likelihood equation $\left.\frac{\partial}{\partial q}\ln p(y;q)\right|_{q=\hat{q}}=0$.
- $\exists$ efficient estimator $\hat{q}$ $\Rightarrow$ distribution of the observation must belong to the exponential family. The efficient estimator can be found by the ML estimator.

CRB

- $\hat{q}$ unbiased estimator of q, then $\mathbb{E}\left[\left(\hat{q}(Y)-\vec{q}\right)\left(\hat{q}(Y)-\vec{q}\right)^T\right]\geq I^{-1}(\vec{q})$ with equality iff $\boxed{\nabla_{\vec{q}}\ln p(y;\vec{q})=I(\vec{q})\left(\hat{q}(y)-\vec{q}\right)}$
- Let $\hat{g}(y)$ be an unbiased estimator of $\vec{g}(\vec{q})$, then $\mathbb{E}\left[\left(\hat{g}(\vec{Y})-\vec{g}(\vec{q})\right)\left(\hat{g}(\vec{Y})-\vec{g}(\vec{q})\right)^T\right]\geq\left(d\vec{g}(\vec{q})\right)I^{-1}(q)\left(d\vec{g}(\vec{q})\right)^T$, with equality iff $\hat{g}(\vec{y})-\vec{g}(\vec{q})=\left(d\vec{g}(\vec{q})\right)I^{-1}(q)\nabla_{\vec{q}}\ln p(y;\vec{q})$.
- Let $\vec{Y}\sim\mathcal{N}\left(\vec{m}_{\vec{q}},\Sigma_{\vec{q}}\right)$. Then, $\left[I(\vec{q})\right]_{ij}=\left(\frac{\partial\vec{m}_{\vec{q}}}{\partial q_i}\right)^T\Sigma_{\vec{q}}^{-1}\left(\frac{\partial\vec{m}_{\vec{q}}}{\partial q_i}\right)+\frac{1}{2}tr\left\{\Sigma_{\vec{q}}^{-1}\frac{\partial\Sigma_{\vec{q}}}{\partial q_i}\Sigma_{\vec{q}}^{-1}\frac{\partial\Sigma_{\vec{q}}}{\partial q_j}\right\}$ where $\frac{\partial\vec{m}_{\vec{q}}}{\partial q_i}=\left[\frac{\partial m_1(\vec{q})}{\partial q_i},\ldots,\frac{\partial m_n(\vec{q})}{\partial q_i}\right]^T$, $\frac{\partial\Sigma_{\vec{q}}}{\partial q_k}=\left[\frac{\partial\left[\Sigma_{\vec{q}}\right]_{ij}}{\partial q_k}\right]$.
  - Let $\vec{Y}\sim\mathcal{N}\left(\vec{m}_{\vec{q}},\Sigma\right)$, then $I(\vec{q})=\left(d\vec{m}_{\vec{q}}\right)^T\Sigma^{-1}d\vec{m}_{\vec{q}}$. Also, $\nabla_{\vec{q}}\ln p(\vec{y};\vec{q})=\left(d\vec{m}_{\vec{q}}\right)^T\Sigma^{-1}\left(\vec{y}-\vec{m}_{\vec{q}}\right)$.

- **Linear model**: $\vec{X} = H\vec{q} + \vec{W}$, $\vec{W} \sim \mathcal{N}(0,\Sigma)$. Then $\vec{X} \sim \mathcal{N}(H\vec{q},\Sigma)$, and $I(\vec{q}) = H^T\Sigma^{-1}H$. $\nabla_{\vec{q}} \ln p(\vec{y};\vec{q}) = H^T\Sigma^{-1}H\left(\left(H^T\Sigma^{-1}H\right)^{-1}H^T\Sigma^{-1}\vec{y} - \vec{q}\right)$. $\boxed{\left(H^T\Sigma^{-1}H\right)^{-1}H^T\Sigma^{-1}\vec{y}}$ is UMVU, efficient, Gaussian, ML, Least-square. Need H full column rank for identifiability. $\hat{q} \sim \mathcal{N}\left(\vec{q},\left(H^T\Sigma^{-1}H\right)^{-1}\right)$.
- Let $\vec{Y} \sim \mathcal{CN}\left(\vec{m}_{\vec{q}},\bar{\Sigma}_{\vec{q}}\right)$, real $\vec{q}$. Then $\left[I(\vec{q})\right]_{ij} = 2\mathrm{Re}\left\{\left(\frac{\partial \vec{m}_{\vec{q}}}{\partial q_i}\right)^H \bar{\Sigma}_{\vec{q}}^{-1}\left(\frac{\partial \vec{m}_{\vec{q}}}{\partial q_i}\right)\right\} + \frac{1}{2}tr\left\{\bar{\Sigma}_{\vec{q}}^{-1}\frac{\partial \bar{\Sigma}_{\vec{q}}}{\partial q_i}\bar{\Sigma}_{\vec{q}}^{-1}\frac{\partial \bar{\Sigma}_{\vec{q}}}{\partial q_j}\right\}$.

State Estimation

- State Estimation: 1) states: $\vec{S}_{n+1} = A_n\vec{S}_n + \vec{U}_n$. 2) observation: $\vec{Y}_n = H_n\vec{S}_n + \vec{W}_n$. Known distribution of $\vec{S}_0$, input sequence $\{\vec{U}_n\}$, observation noise $\{\vec{W}_n\}$. $E\left[\vec{S}_0\right] = \vec{s}_{0|-1}$, $VAR\left[\vec{S}_0\right] = \Sigma_{0|-1}$. Find the MMSE estimator of $\vec{S}_n$ given $\vec{Y}_n,\vec{Y}_{n-1},\ldots$, i.e., $\hat{s}_{n|n} = \mathbb{E}\left[\vec{S}_n\middle|\vec{y}_n,\vec{y}_{n-1},\ldots\right]$.

Discrete-Time Kalman–Bucy

- $X_{n+1} = F_nX_n + G_nU_n$, $Y_n = H_nX_n + V_n$. $Q_t = Cov(U_t)$, $R_t = Cov(V_t)$. ½ $\hat{X}_{0|-1} = E[X_0]$, $\Sigma_{0|-1} = \Sigma_0 = Cov(X_0)$ ½ $\Sigma_{t|t-1} = Cov\left(\vec{X}_t\middle|\vec{Y}_0^{t-1}\right)$ ½ Kalman gain matrix $K_t = \Sigma_{t|t-1}H_t^H\left(H_t\Sigma_{t|t-1}H_t^H + R_t\right)^{-1}$. ½ $\hat{X}_{t|t} = \mathbb{E}\left[\vec{X}_t\middle|\vec{Y}_0^t\right] = \hat{X}_{t|t-1} + K_t\left(Y_t - H_t\hat{X}_{t|t-1}\right)$.½ $\Sigma_{t|t} = \Sigma_{t|t-1} - K_tH_t\Sigma_{t|t-1}$.½ $\hat{X}_{t+1|t} = \mathbb{E}\left[\vec{X}_{t+1}\middle|\vec{Y}_0^t\right] = F_t\hat{X}_{t|t}$. ½ $\Sigma_{t+1|t} = F_t\Sigma_{t|t}F_t^T + G_tQ_tG_t^T$.

Kalman Filtering

- Notation: $\vec{y}_{-\infty}^t = \{\vec{y}_t,\vec{y}_{t-1},\ldots\}$. $\hat{s}_{t|t-1} = E\left[\vec{S}_t\middle|\vec{y}_{-\infty}^{t-1}\right]$ = the MMSE prediction of $\vec{S}_t$ from the past samples. $\Sigma_{t|t-1} = E\left[\left(\vec{S}_t - \hat{s}_{t|t-1}\right)\left(\vec{S}_t - \hat{s}_{t|t-1}\right)^H\middle|\vec{y}_{-\infty}^t\right]$. $\hat{s}_{t|t} = E\left[\vec{S}_t\middle|\vec{y}_{-\infty}^t\right]$ (the MMSE filter.) $\Sigma_{t|t} = \mathbb{E}\left[\left(\vec{S}_t - \hat{s}_{t|t}\right)\left(\vec{S}_t - \hat{s}_{t|t}\right)^H\middle|\vec{y}_{-\infty}^t\right]$.
- Gaussian Model: $\{\vec{u}_n\}$, $\{\vec{w}_n\}$ are zero mean, independent, <u>Gaussian</u>. $\Lambda_{U_n} = E[\vec{u}_n\vec{u}_n']$. $\Lambda_{W_n} = E[\vec{w}_n\vec{w}_n']$. $\vec{s}_0 \sim N(\bar{s}_0,\Lambda_0)$ independent of $\{\vec{u}_n\}$, $\{\vec{w}_n\}$.
  - Initialization: $\hat{s}_{0|-1} = E\left[\vec{S}_0\right]$, $\Sigma_{0|-1} = VAR\left[\vec{S}_0\right] = \Sigma_{\vec{S}_0\vec{S}_0}$, $\hat{y}_{0|-1} = H_0\hat{s}_{0|-1}$.
  - Measurement Update: filtering: $K_k = \Sigma_{k|k-1}H_k'\left(H_k\Sigma_{k|k-1}H_k' + \Sigma_{\vec{W}_k}\right)^{-1}$ ½ $\hat{s}_{k|k} = \hat{s}_{k|k-1} + K_k\left(\vec{y}_k - \hat{y}_{k|k-1}\right)$ ½ $\Sigma_{k|k} = \Sigma_{k|k-1} - K_kH_k\Sigma_{k|k-1}$

- Time Update: prediction: $\hat{s}_{k+1|k} = A_k \hat{s}_{k|k}$ ½ $\Sigma_{k+1|k} = A_k \Sigma_{k|k} A_k' + \Sigma_{\bar{U}_k}$ ½
  $\hat{y}_{k+1|k} = H_{k+1} \hat{s}_{k+1|k}$
- Same formula for linear MMSE of non Gaussian.

Example

- Normal i.i.d.: $\hat{m}_{ML} = \frac{1}{n}\sum y_k$; $\hat{s}^2_{ML} = \frac{1}{n}\sum (y_k - \hat{m}_{ML})^2$ biased.
- Exponential i.i.d.: $p(\bar{y};q) = \prod_{i=1}^{n} q e^{-q y_i} I(y_i > 0) = q^n e^{-q\sum_{i=1}^{n} y_i} \left(\prod_{i=1}^{n} I(y_i > 0) \right)$.
  $\hat{q}(\bar{y}) = \frac{n-1}{\sum_{i=1}^{n} y_i}$ is UMVU. $\hat{q}_{ML}(\bar{y}) = \frac{n}{\sum_{i=1}^{n} y_i}$ biased.
- $X_i \overset{i.i.d.}{\sim} \mathscr{P}(l)$. $p(\bar{x};l) = \frac{l^{\sum_{i=1}^{n} x_i} e^{-nl}}{\prod_{j=1}^{n}(x_j!)} = l^{\sum_{i=1}^{n} x_i} e^{-nl} \frac{1}{\prod_{j=1}^{n}(x_j!)} = \left(\sum_{i=1}^{n} x_i ; l \right) h(\bar{x})$.
- $X_i \overset{i.i.d}{\sim} \mathcal{U}(0,q)$. $p(\bar{x};q) = \underbrace{\frac{1}{q^n} I(\max\{x_i\} < q)}_{g(\max\{x_i\},q)} I(\min\{x_i\} > 0)$. $t(\bar{x}) = \max\{x_i\}$ is complete and sufficient. $\hat{q}(\bar{y}) = \frac{n+1}{n}\max\{y_i\}$ is UMVU.
- Binary i.i.d.: $p(\bar{x};q) = q^{\sum_{k=1}^{n} y_k} (1-q)^{n-\sum_{k=1}^{n} y_k}$. $\sum_{k=1}^{n} y_k$ is complete and sufficient.
  $\hat{q}_{ML} = \hat{q}_{UMVU} = \frac{1}{n}\sum_{k=1}^{n} y_k$.
- Binomial: $p(y;q) = \binom{n}{y} q^y (1-q)^{n-y}$; y is complete. No unbiased estimator for $g(q) = \frac{1}{q}$.

ML Estimator

- The ML estimator of parameter $\bar{q}$ from $\bar{Y} \sim p(\bar{y};\bar{q})$ $q \in \Theta$ is $\hat{q}_{ML}(\bar{y})$ =
  $\boxed{\operatorname{argmax}_{q\in\Theta} p(\bar{y};\bar{q})} = \boxed{\operatorname{argmax}_{q\in\Theta} \ln p(\bar{y};\bar{q})}$.
- The best linear unbiased estimator (BLUE) is $\hat{q}_{BLUE} = A_{BLUE}\bar{y}$ where
  $A_{BLUE} = \operatorname{argmin} \mathbb{E}\left[\|q - AY\|^2\right]$ subject to $\mathbb{E}[AY] = q$.
- For linear model $X = Hq + W$ with zero mean noise, $\hat{q}_{BLUE} = \hat{q}_{ML}$.

- For the K-parameter exponential family, let $\mathcal{C}$ be the interior of the range of $\left\{\left(c_1(q),\ldots,c_K(q)\right)^T, q \in \Lambda\right\}$. If $\mathbb{E}\left[t_i(Y)\right] = t_i(y)$, i = 1, …, K have a solution $\hat{q}(y)$ for which $\left(c_1(\hat{q}),\ldots,c_K(\hat{q})\right)^T \in \mathcal{C}$, then $\hat{q}$ is the unique ML estimator of q.
- **Invariance**: Let $g(q): \Theta \xrightarrow{onto} \Phi$, $g^{-1}(f): \Phi \longrightarrow \{A: A \subset \Theta\}$ be the inverse image. Define $\ell(y;f) \triangleq \sup_{q \in g^{-1}(f)} p(y;q)$. If $\hat{q}_{ML}$ is the ML estimate of q, then $\hat{f}_{ML} \triangleq \operatorname{argsup}_{f\in\Phi} \ell(y;f) = g\left(\hat{q}_{ML}\right)$.
- $D\left(q_0 \| q\right) \triangleq \mathbb{E}_{q_0}\left[\ln \frac{p(Y;q_0)}{p(Y;q)}\right] = \int_y p(y;q_0) \ln \frac{p(y;q_0)}{p(y;q)} dy \geq 0$ with equality iff $p(y;q_0) = p(y;q)$ a.e. If q_0 is identifiable, then $D\left(q_0 \| q\right) = 0 \Leftrightarrow q = q_0$.
- q_0 is the global minimum of $D\left(q_0 \| q\right)$, and $\min_{q\in\Theta} D\left(q_0 \| q\right) \Leftrightarrow \max_{q\in\Theta} \mathbb{E}_{q_0}\left[\ln p(Y;q)\right]$.
- For i.i.d. Y_i, $\hat{q}_{ML} = \operatorname{argmax}_q \frac{1}{N}\sum_{i=1}^{N} \ln p(y_i;q) \xrightarrow[n\to\infty]{} \operatorname{argmax}_q \mathbb{E}_{q_0}\left[\ln p(Y;q)\right]$.
- To solve for ML: $s\left(\bar{y};\bar{q}\right) = \nabla_{\bar{q}} \ln p\left(\bar{y};\bar{q}\right)\Big|_{\bar{q}=\hat{q}_{ML}} = 0$.
  - Newton-Raphson: $q^{(k+1)} = q^{(k)} - \left(J^{-1}\left(y;q^{(k)}\right)\right)\left(s\left(y;q^{(k)}\right)\right)$.
    $J(y;q) = \nabla_{\bar{q}}^2 \ln p\left(\bar{y};\bar{q}\right)$.
  - Scoring Method: $q^{(k+1)} = q^{(k)} + I^{-1}\left(q^{(k)}\right)\left(s\left(y;q^{(k)}\right)\right)$.
- Let the complete data $Z = \begin{bmatrix} S \\ Y \end{bmatrix} \sim p(z;q)$. Only $Y \sim p(y;q)$ is observed.
- $Q\left(q^{(2)},q^{(1)}\right) > Q\left(q^{(1)},q^{(1)}\right) \Rightarrow \ln p\left(y;q^{(2)}\right) \geq \ln p\left(y;q^{(1)}\right)$.
- EM: $Q\left(q;q^{(k)}\right) = \mathbb{E}_{q^{(k)}}\left[\ln p(Z;q) \middle| Y = y\right]$, $q^{(k+1)} = \operatorname{argmax}_q Q\left(q;q^{(k)}\right)$.
- If distribution of S does not depend on q,
  $Q\left(q;q^{(k)}\right) = \mathbb{E}_{q^{(k)}}\left[\ln p\left(Y|S;q\right) \middle| Y = y\right] + \text{constant}$
- Asymptotically unbiased ≡ $\lim_{n\to\infty} \mathbb{E}\left(\hat{q}(Y)\right) - q = 0$
- Consistency: (d) distribution $\lim_{n\to\infty} p_{\hat{q}}(q) = p_q(q)$, (p) weak $\lim_{n\to\infty} \Pr\left[\left|\hat{q}(Y) - q\right| > d\right] = 0 \ \forall q$, (w.p.1) strong $\Pr\left[\lim_{n\to\infty} \hat{q}(Y) = q\right] = 1 \ \forall q$, (m.s.) mean square $\lim_{n\to\infty} \mathbb{E}\left[\left\|\hat{q}(Y) - q\right\|^2\right] = 0$. (w.p.1) ⇒ (p) ⇒ (d). (m.s.) ⇒ p. p and bounded Θ ⇒ (m.s.)
- Asymptotically Normal: $\sqrt{n}\left(\hat{q} - q\right) \rightarrow\sim N\left(0, \Sigma(q)\right)$.

- Asymptotically efficient (BAN: best asymptotically normal) ≡ $\lim_{n\to\infty} \frac{Var\left(\hat{q}(Y)\right)}{CRB(q)} = 1$. ≡ $\sqrt{n}\left(\hat{q} - q\right) \to\sim N\left(0, I_0^{-1}\right)$; $I_0(q) = \lim_{n\to\infty} \frac{1}{n} I(q)$.
- Y_i i.id. Under regularity conditions, 1) $\hat{q}_{ML}^{(n)} \to q$ (weak). 2) $\hat{q}_{ML}^{(n)}$ is asymptotically Gaussian and asymptotically efficient.

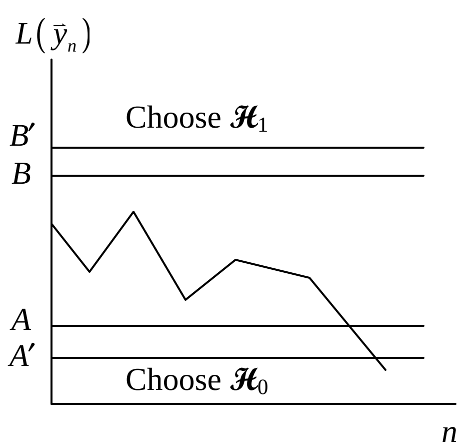
Sequential Detection

- Fixed sample size (FSS) detector. Example: Consider the n-sample simple binary hypotheses $\mathcal{H}_0$ vs. $\mathcal{H}_1$: $\mathcal{H}_i = Y_k \overset{iid}{\sim} \mathcal{N}\left(m_i, s^2\right)$, i = 0, 1, k = 1, 2, …, $m_1 > m_0$. Then $\ln L\left(\vec{Y}_n\right) = \left(\frac{m_1 - m_0}{s^2}\right)\left(\sum_{k=1}^{n} Y_k\right) - n\frac{m_1^2 - m_0^2}{2s^2}$.
$$\ln L\left(\vec{Y}_n\right) \sim \begin{cases} \mathcal{N}\left(n\frac{(m_1 - m_0)^2}{2s^2}, n\frac{(m_1 - m_0)^2}{s^2}\right), & \mathcal{H}_1 \\ \mathcal{N}\left(-n\frac{(m_1 - m_0)^2}{2s^2}, n\frac{(m_1 - m_0)^2}{s^2}\right), & \mathcal{H}_0 \end{cases}$$
. The minimum n such that the optimal detector has size < a, and power > b is $n = \left(\frac{s}{m_1 - m_0}\left(Q^{-1}(a) - Q^{-1}(b)\right)\right)^2$.
- Simple binary hypotheses $\mathcal{H}_0$ vs. $\mathcal{H}_1$. $\mathcal{H}_i = Y_k \overset{iid}{\sim} p\left(y; q_i\right)$, k = 1, 2, …
- A sequential detector (f, d) is defined by 1) stopping rule sequence $\left[f_n\right]$: $f_n\left(\vec{y}_n\right) = \begin{cases} 1 \equiv \text{ stop data collection \& make decision} \\ 0 \equiv \text{ continue data collection} \end{cases} : \mathbb{R}^n \to \{0,1\}$ .terminal decision rule sequence $\left[d_n\right]$: $d_n\left(\vec{y}_n\right) = \Pr\left[D = 1 \middle| \vec{Y}_n = \vec{y}_n\right]$. 2) Stop time: $N(f) = \min\left\{k : f_k\left(\vec{Y}_k\right) = 1\right\}$.
- The sequential probability ratio test:

SPRT(*A*,*B*): $L\left(\vec{y}_n\right) = \frac{p\left(\vec{y}_n; q_1\right)}{p\left(\vec{y}_n; q_0\right)}$,

$$f_n\left(\vec{y}_n\right) = \begin{cases} 1 \equiv \text{ stop}, & L\left(\vec{y}_n\right) \geq B, \\ & \text{or } L\left(\vec{y}_n\right) \leq A \\ 0, & L\left(\vec{y}_n\right) \in (A, B) \end{cases},$$

$$d_n\left(\vec{y}_n\right) = \begin{cases} 1, & L\left(\vec{y}_n\right) \geq B \\ 0, & L\left(\vec{y}_n\right) \leq A \end{cases}$$



- SPRT is optimal in the Bayesian problem. SPRT satisfies $A \ge \frac{1-b}{1-a}$ and $B \le \frac{b}{a}$.
- The **Wald-Wolfowitz Theorem**: The SPRT(A,B) detector (f_*,d_*) has the minimum stop time among all detectors (including FSS detector) with size no larger and power no less than those of (f_*,d_*). $\left.\begin{matrix} P_F(d) \le P_F(d_*) \\ P_D(d) \ge P_D(d_*) \end{matrix}\right\} \Rightarrow \mathbb{E}_{q_i}\left[N(f)\right] \ge \mathbb{E}_{q_i}\left[N(f_*)\right]$, $i = 0,1$
- $Z_i = \ln \frac{p(Y_i;q_1)}{p(Y_i;q_0)}$. $z_i = \ln \frac{p(y_i;q_1)}{p(y_i;q_0)}$. $\ln L(\vec{y}_n) = \sum_{i=1}^{n} z_i$.
- **Wald's Approximations** : Given a and b, the optimal SPRT(A,B) can be approximated by $\text{SPRT}(A',B')$ with $\boxed{A' = \frac{1-b}{1-a}}$ and $\boxed{B' = \frac{b}{a}}$.
  - $\frac{1-b'}{1-a'} \le \frac{1-b}{1-a} = A' \le A \overset{?}{<} B \le B' = \frac{b}{a} \le \frac{b'}{a'} \Rightarrow a < b$.
  - $(A,B) \subset (A',B')$ $\Rightarrow$ the approximation requires more samples.
  - $P_F(d_*') + P_M(d_*') = a' + (1-b') \le a + (1-b) = P_F(d_*) + P_M(d_*)$
- **The Wald's Equation**: Let Z_i be independent and $i.i.d.$ with $EZ \le \infty$. Let N be a stopping time, then, $\mathbb{E}[Z_1 + \cdots + Z_N] = \mathbb{E}[Z]\mathbb{E}[N]$.
- $$\boxed{\begin{aligned} \mathbb{E}_{q_0}[N] &\approx \frac{a \ln B + (1-a)\ln A}{\mathbb{E}_{q_0}[Z]} \approx \frac{a \ln \frac{b}{a} + (1-a)\ln \frac{1-b}{1-a}}{\mathbb{E}_{q_0}[Z]} \\ \mathbb{E}_{q_1}[N] &\approx \frac{b \ln B + (1-b)\ln A}{\mathbb{E}_{q_1}[Z]} \approx \frac{b \ln \frac{b}{a} + (1-b)\ln \frac{1-b}{1-a}}{\mathbb{E}_{q_1}[Z]} \end{aligned}}$$

  $\mathbb{E}_{q_0}[Z] = \int_y p(y;q_0) \ln \frac{p(y;q_1)}{p(y;q_0)} dy$. $\mathbb{E}_{q_1}[Z] = \int_y p(y;q_1) \ln \frac{p(y;q_1)}{p(y;q_0)} dy$.

Review

- $\sum_{k=0}^{n} k = \frac{n(n+1)}{2}$, $\sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$, $f(x) = \int_{a(x)}^{b(x)} g(x,y)dy$, $f'(x) =$ $b'(x)g(x,b(x)) - a'(x)g(x,a(x)) + \int_{a(x)}^{b(x)} \frac{\partial g}{\partial x}(x,y)dy$,. $\int_0^\infty x^n e^{-ax} dx = \frac{\Gamma(n+1)}{a^{n+1}}$, for $n \in \mathbb{N}$, $\overset{n\in\mathbb{N}}{=} \frac{n!}{a^{n+1}}$. $\int_0^\infty e^{-ax^2} dx = \frac{1}{2}\sqrt{\frac{p}{a}}$. $\int_{-\infty}^{\infty} e^{-(ax^2+bx+c)} dx = \sqrt{\frac{p}{a}} e^{\frac{b^2-4ac}{4a}}$.
  $\int x e^{-l x} dx = -\frac{1}{l^2}(1 + l x)e^{-l x}$

- $\boxed{\nabla_{\vec{x}}\left(\vec{f}^T(\vec{x})\right)=\left(d\vec{f}(\vec{x})\right)^T}$. $df(y)=\left(\frac{\partial f}{\partial x_1}(y),\ldots,\frac{\partial f}{\partial x_n}(y)\right)$. $d\left(\|\vec{x}\|^2\right)=2\vec{x}^T$.
  $d\left(A\vec{x}+\vec{b}\right)=A$. $d\left(\vec{a}^T\vec{x}\right)=\vec{a}^T$. $\boxed{d\left(\vec{f}^T(\vec{x})\vec{g}(\vec{x})\right)=\vec{f}^T(\vec{x})d\vec{g}(\vec{x})+\vec{g}^T(\vec{x})d\vec{f}(\vec{x})}$.
  $d\left(\vec{f}^T(\vec{x})Q\vec{f}(\vec{x})\right)=2\vec{f}^T(\vec{x})Qd\vec{f}(\vec{x})$. $d\left(\left\|\vec{f}(\vec{x})\right\|^2\right)=2\vec{f}^T(\vec{x})d\vec{f}(\vec{x})$.
- $\nabla_{\vec{x}}\|\vec{x}\|^2=2\vec{x}$, $\nabla_{\vec{x}}\left(A\vec{x}+\vec{b}\right)=A^T$, $\nabla_{\vec{x}}\left(\vec{a}^T\vec{x}\right)=\vec{a}$, $\nabla_{\vec{x}}\left(\vec{f}^T(\vec{x})\vec{g}(\vec{x})\right)=\nabla_{\vec{x}}\left(\vec{g}^T(\vec{x})\right)\vec{f}(\vec{x})$ $+\nabla_{\vec{x}}\left(\vec{f}^T(\vec{x})\right)\vec{g}(\vec{x})$. $\nabla_{\vec{x}}\left(\vec{f}^T(\vec{x})Q\vec{f}(\vec{x})\right)=2\nabla_{\vec{x}}\left(\vec{f}^T(\vec{x})\right)Q\vec{f}(\vec{x})$. $\nabla_{\vec{x}}\left(\vec{x}^TQ\vec{x}\right)=2Q\vec{x}$.
  $\nabla_{\vec{x}}\left(\left\|\vec{f}(\vec{x})\right\|^2\right)=2\nabla_{\vec{x}}\left(\vec{f}^T(\vec{x})\right)\vec{f}(\vec{x})$.
- $\mathbb{E}\left[f(X,Y)\middle|Y=y\right]=\int f(x,y)\,p_{X|Y}(x|y)dx$ a constant. $\mathbb{E}\left[f(X,Y)\middle|Y\right]$ is a r.v.
- $\mathbb{E}\left[f(X)g(Y)\right]=\mathbb{E}\left[g(Y)\mathbb{E}\left[f(X)\middle|Y\right]\right]$. Set $g(y)\equiv 1$, $f(x)=x$, then have
  $\mathbb{E}[X]=\mathbb{E}\left[\mathbb{E}\left[X\middle|Y\right]\right]$.
- $\vec{Z}=\begin{pmatrix}\vec{X}\\ \vec{Y}\end{pmatrix}$ is jointly Gaussian. $\begin{pmatrix}\vec{X}\\ \vec{Y}\end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix}\vec{m}_{\vec{X}}\\ \vec{m}_{\vec{Y}}\end{pmatrix},\begin{pmatrix}\Lambda_{\vec{X}\vec{X}} & \Lambda_{\vec{X}\vec{Y}}\\ \Lambda_{\vec{Y}\vec{X}} & \Lambda_{\vec{Y}\vec{Y}}\end{pmatrix}\right)$. Then,
  $p(x|y)\sim\mathcal{N}\left(E\left[X\middle|y\right],\Lambda_{X|y}\right)$ where $E\left[X\middle|y\right]=\vec{m}_{\vec{X}}+\Lambda_{\vec{X}\vec{Y}}\Lambda_{\vec{Y}\vec{Y}}^{-1}\left(\vec{y}-\vec{m}_{\vec{Y}}\right)$,
  $\Lambda_{X|y}=\Lambda_{\vec{X}\vec{X}}-\Lambda_{\vec{X}\vec{Y}}\Lambda_{\vec{Y}\vec{Y}}^{-1}\Lambda_{\vec{Y}\vec{X}}$.
  Define $K=\Lambda_{\vec{X}\vec{Y}}\Lambda_{\vec{Y}}^{-1}$. Then $E\left[X\middle|y\right]=\vec{m}_{\vec{X}}+K\left(\vec{y}-\vec{m}_{\vec{Y}}\right)$, $\Lambda_{X|y}=\Lambda_{\vec{X}\vec{X}}-K\Lambda_{\vec{Y}\vec{X}}$.
- $\left(\vec{X},\vec{Y},\vec{W}\right)$ jointly Gaussian. $\vec{W}\perp\!\!\!\perp\left(\vec{X},\vec{Y}\right)$, $\vec{V}=B\vec{X}+\vec{W}$. Then
  $\vec{V}|\vec{y}\sim\mathcal{N}\left(\mathbb{E}\left[\vec{V}\middle|\vec{y}\right],\Lambda_{\vec{V}|\vec{y}}\right)$ where $\mathbb{E}\left[\vec{V}\middle|y\right]=B\mathbb{E}\left[\vec{X}\middle|y\right]+\mathbb{E}\vec{W}$,
  and $\Lambda_{\vec{V}|y}=B\Lambda_{\vec{X}|y}B^T+\Lambda_{\vec{W}\vec{W}}$.
- $\left(\vec{a}\vec{a}^T+cI_{n\times n}\right)^{-1}=\frac{1}{c}I-\frac{1}{c\left(\vec{a}^T\vec{a}+c\right)}\vec{a}\vec{a}^T$.
- Δ_{11} = **Schur complement** of A_{11} = $A_{22}-A_{21}A_{11}^{-1}A_{12}$ ½ $\Delta_{22}=A_{11}-A_{12}A_{22}^{-1}A_{21}$.
  $$\underbrace{\begin{bmatrix}I & 0\\ -A_{21}A_{11}^{-1} & I\end{bmatrix}}_{T}\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\underbrace{\begin{bmatrix}I & -A_{11}^{-1}A_{12}\\ 0 & I\end{bmatrix}}_{T'}=\begin{bmatrix}A_{11} & 0\\ 0 & \Delta_{11}\end{bmatrix};$$
- $A^{-1}=\begin{bmatrix}\Delta_{22}^{-1} & -\Delta_{22}^{-1}A_{12}A_{22}^{-1}\\ -\Delta_{11}^{-1}A_{21}A_{11}^{-1} & \Delta_{11}^{-1}\end{bmatrix}$ | $\left(A+BCD\right)^{-1}=A^{-1}-A^{-1}B\left(C^{-1}+DA^{-1}B\right)^{-1}DA^{-1}$
  $\left(A+BCD\right)^{-1}=A^{-1}-A^{-1}B\left(I+CDA^{-1}B\right)^{-1}CDA^{-1}$
- $\left(A=A'\right)$ and $\left(A\geq 0\right)\Rightarrow TAT'\geq 0\Rightarrow\Delta_{11}\geq 0$.

- $\mathcal{N}(m,s^2)$: $f_X(x)=\frac{1}{\sqrt{2p}\,s}e^{-\frac{(x-m)^2}{2s^2}}$, $E\left[e^{-jvX}\right]=e^{jmv-\frac{1}{2}v^2s^2}$. $\mathcal{N}\left(\bar{m}_{\bar{q}},\bar{\Sigma}_{\bar{q}}\right)$

  $\frac{1}{(2p)^{\frac{n}{2}}\sqrt{\det(\Lambda)}}e^{-\frac{1}{2}(\underline{x}-\underline{m})^T\Lambda^{-1}(\underline{x}-\underline{m})}$; i.i.d. $\frac{1}{\left(2ps^2\right)^{\frac{n}{2}}}\exp\left\{-\frac{\|x_i-m\|^2}{2s^2}\right\}=$

  $\frac{1}{\left(2ps^2\right)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2s^2}\sum_{i=1}^{n}x_i^2+\frac{m}{s^2}\sum_{i=1}^{n}x_i-\frac{m^2n}{2s^2}\right\}$. ; $\mathcal{CN}\left(\bar{m}_{\bar{q}},\bar{\Sigma}_{\bar{q}}\right)$.

  $\frac{1}{p^n\det(\Lambda)}e^{-(\underline{x}-\underline{m})^H\Lambda^{-1}(\underline{x}-\underline{m})}$.

- $Q(0)=\frac{1}{2}$, $Q(-z)=1-Q(z)$ ô $Q^{-1}(1-Q(z))=-z$ ô $\boxed{P[X>x]=Q\left(\frac{x-m}{s}\right)}$

  ô $P[X<x]=1-Q\left(\frac{x-m}{s}\right)=Q\left(-\frac{x-m}{s}\right)$

- **Poisson** $\mathcal{P}(l)$, $e^{-l}\frac{l^i}{i!}$; Ω = N, 0 ≤ λ, EX = l, VAR(X) = l, $\Phi_X(u)=Ee^{iuX}$ = $e^{l\left(e^{iu}-1\right)}$

  ô **Binomial** $\binom{n}{k}p^k(1-p)^{n-k}$; np, $np(1-p)$, $(pe^{iu}+1-p)^n$ ô **Uniform** $\mathcal{U}(a,b)$, $\frac{a+b}{2}$,

  $\frac{(b-a)^2}{12}$, $e^{iu\frac{b+a}{2}}\frac{\sin\left(u\frac{b-a}{2}\right)}{u\frac{b-a}{2}}$ ô **Exponential** $\mathcal{E}(a)$, $\frac{1}{a}$, $\frac{1}{a^2}$, $\frac{a}{a-iu}$. ô **Laplacian**

  $\mathcal{L}(\alpha)$, $\frac{a}{2}e^{-a|x|}$;α > 0, $\begin{cases}\frac{1}{2}e^{ax} & x<0\\ 1-\frac{1}{2}e^{-ax} & x\geq 0\end{cases}$, 0, $\frac{2}{a^2}$, $\frac{a^2}{a^2+u^2}$.

- Gamma function: $\Gamma(q)=\int_0^\infty x^{q-1}e^{-x}dx$. ; $q>0$. $\Gamma(n)=(n-1)!$ for $n\in\mathbb{N}$. $\Gamma(n+1)=n!$

  if $n\in\mathbb{N}\cup\{0\}$. 0! = 1. $\Gamma\left(\frac{1}{2}\right)=\sqrt{p}$. $\Gamma(x+1)=x\Gamma(x)$. **Gamma distribution**: $\Gamma(q,l)$. $p(x)=\frac{l^q x^{q-1}e^{-lx}}{\Gamma(q)}$. $q>0$. $x\geq 0$. Exponential distribution

  $\mathcal{E}(l)$ is $\Gamma(1,l)$.

- Beta distribution: $p(z)=b_{q_1,q_2}(z)=\frac{\Gamma(q_1+q_2)}{\Gamma(q_1)\Gamma(q_2)}z^{q_1-1}(1-z)^{q_2-1}$; $z\in(0,1)$

- Let $X_i \sim p(x_i) = \dfrac{l^{q_i} x^{q_i - 1} e^{-l x_i}}{\Gamma(q_i)}$, independent. Then $Z_1 = \dfrac{X_1}{X_1 + X_2}$ and $Z_2 = X_1 + X_2$ are independent. $Z_1 = \dfrac{X_1}{X_1 + X_2} \sim b_{q_1, q_2}(z)$. $Z_2 = X_1 + X_2 \sim \Gamma(q_1 + q_2, l)$. $\sum_i X_i \sim \Gamma\left(\sum_i q_i, l\right)$.
- **Central chi-square distribution**: $X \sim \mathcal{N}(0, s^2)$. $Y = X^2$. Then $p(y) = \dfrac{1}{\sqrt{2p y s}} e^{-\frac{y}{2s^2}}, y \geq 0$. $\Phi(u) = \dfrac{1}{(1 - j2us^2)^{\frac{1}{2}}}$. **chi-square** (or gamma): $X_i \overset{i.i.d.}{\sim} \mathcal{N}(0, s^2)$. $Y = \sum_{i=1}^{n} X_i^2$. Then $p(y) = \dfrac{1}{(2s^2)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2} - 1} e^{-\frac{y}{2s^2}} = \Gamma\left(\dfrac{n}{2}, \dfrac{1}{2s^2}\right), y \geq 0$. $\Phi(u) = (1 - j2us^2)^{-\frac{n}{2}}$. $\mathbb{E}[Y] = ns^2$, $Var[Y] = 2ns^4$.

Decorrelation

- If $A \geq 0$ is the covariance matrix $\left(E[\vec{x}\vec{x}']\right)$ of a zero mean random vector $\vec{x} = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \end{bmatrix}$. The vector $\vec{x}$ can be decorrelated via transform $\vec{y} = \begin{bmatrix} \vec{y}_1 \\ \vec{y}_2 \end{bmatrix} = T\vec{x} = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 - A_{21} A_{11}^{-1} \vec{x}_1 \end{bmatrix}$ with covariance $Cov(\vec{y}) = E[\vec{y}\vec{y}'] = \begin{bmatrix} A_{11} & 0 \\ 0 & \Delta_{11} \end{bmatrix}$. $\Delta_{11} = A_{22} - A_{21} A_{11}^{-1} A_{12} = Cov\left(\vec{x}_2 - A_{21} A_{11}^{-1} \vec{x}_1\right) \geq 0$ with equality iff $\Pr\left[\vec{x}_2 = A_{21} A_{11}^{-1} \vec{x}_1\right] = 1$.