A machine processes parts one at a time. Processing times are iid $\sim \mathcal{U}(1,5)$.

Approximate probability that # parts processed within 320 time units, N_{320}, is at least 100.

No obvious way to express N_{320} as sum of independent variates.

Instead, let X_i be processing time of ith part, $S_{100} = X_1 + \dots + X_{100}$.

The event $\{N_{320} \geq 100\}$ same as $\{S_{100} \leq 320\}$.

Note that $\mu = E[X_i] = 3$, $\sigma^2 = \text{var}(X_i) = \frac{16}{12} = \frac{4}{3}$.

Normalized variable:

$$z = \frac{320 - n\mu}{\sigma\sqrt{n}} = \frac{320 - 300}{\sqrt{100 \cdot \frac{4}{3}}} = 1.73$$

Approx:

$$P(S_{100} \leq 320) \approx \Phi(1.73) = 0.9582.$$

Maximum Likelihood Estimation in discrete case carries over to continuous case: $\hat{\theta}_{ML}(u)$ is value of θ that maximizes $f_\theta(u)$, w.r.t. θ.

Ex Suppose a random variable T has exponential distribution with unknown parameter λ, and we have an observation of $T=t$. Find $\hat{\lambda}_{ML}(t)$.

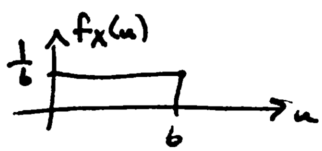
maximize $\lambda e^{-\lambda t}$ w.r.t. λ with t fixed.

$$\frac{d}{d\lambda} \lambda e^{-\lambda t} = (1-\lambda t) e^{-\lambda t}$$

This increases in λ for $0 \le \lambda \le \frac{1}{t}$ and decreases in λ for $\lambda \ge \frac{1}{t}$, so likelihood maximized for $\lambda = \frac{1}{t}$, i.e. $\hat{\lambda}_{ML}(t) = \frac{1}{t}$. In a sense just the inverse sample mean.

Ex Suppose X drawn from uniform distribution on $[0,b]$ where b is unknown parameter. Find "alphabet size" estimator for b, given $X=u$ is observed: $\hat{b}_{ML}(u)$.

$f_b(u)$ is



need to think of it as function of b.

- zero if $b < u$
- jumps to $\frac{1}{u}$ at $b=u$
- as b increases beyond u, function decreases

so max at $b=u$, i.e. $\hat{b}_{ML}(u) = u$.

Maximum Likelihood Estimation in discrete case carries over to continuous case: $\hat{\theta}_{ML}(x)$ is value of θ that maximizes $f_\theta(x)$, w.r.t. θ.

Ex 1. Suppose a random variable T has exponential distribution with unknown parameter λ, and we have an observation of $T=t$. Find $\hat{\lambda}_{ML}(t)$.

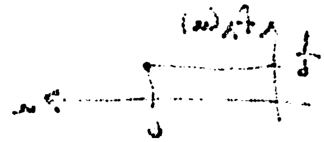
maximize $\lambda e^{-\lambda t}$ w.r.t. λ with t fixed.

$$\frac{d}{d\lambda}\lambda e^{-\lambda t} = (1-\lambda t)e^{-\lambda t}$$

This increases in λ for $0 \le \lambda \le \frac{1}{t}$ and decreases in λ for $\lambda \ge \frac{1}{t}$, so likelihood maximized at $\lambda = \frac{1}{t}$, i.e. $\hat{\lambda}_{ML}(t) = \frac{1}{t}$. If n obs, $\frac{1}{\bar{t}}$, i.e. the inverse sample mean.

Ex 2. Suppose X drawn from uniform distribution on $[0,b]$ where b is unknown parameter. Find "alphabetical" estimator for b, given $X=x$ is observed: $\hat{b}_{ML}(x)$.

$f_b(x)$ is



Now think of it as function of b.

Zero if $b < x$
Jumps to $\frac{1}{x}$ at $b = x$
As b increases beyond x, function decreases.

So max at $b = x$, i.e. $\hat{b}_{ML}(x) = x$.

Hypothesis testing in discrete case also carries over to continuous case, when using likelihood ratio test. Doing the matrices doesn't work.

As before:

$$\Lambda(u) = \frac{f_1(u)}{f_0(u)}$$

LRT:

$$\Lambda(x) \begin{cases} \geq \tau & \text{declare } H_1 \\ < \tau & \text{declare } H_0 \end{cases}$$

under ML, $\tau = 1$.

under MAP, $\tau = \frac{\pi_0}{\pi_1}$

Same notions of error probability as before, MAP minimizes error probability.

Ex Suppose under hypothesis H_i, observation X has $N(m_i, \sigma^2)$, for $i = 0, 1$. where parameters are known, $\sigma^2 > 0$, $m_0 < m_1$. Priors π_0, π_1 known. Find ML, MAP rules and associated error probabilities: p_F, p_m.

The likelihoods are

$$f_i(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(u-m_i)^2}{2\sigma^2} \right\}$$

so $\Lambda(u) = f_1(u)/f_0(u)$

$$= \exp\left\{ -\frac{(u-m_1)^2}{2\sigma^2} + \frac{(u-m_0)^2}{2\sigma^2} \right\} = \exp\left\{ \left(u - \frac{m_0+m_1}{2}\right)\left(\frac{m_1-m_0}{\sigma^2}\right) \right\}.$$

Notice that $\Lambda(X) > 1$ if and only if $X \geq \frac{m_0+m_1}{2}$, so ML rule is

$$X \begin{cases} > \frac{m_0+m_1}{2} & \text{declare } H_1 \\ < \frac{m_0+m_1}{2} & \text{declare } H_0 \end{cases}$$

(threshold test for X rather than for Λ).

hypothesis testing in different cases, also, constructing [illegible] or confidence sets

using likelihood ratio test, Using the Neyman-Pearson lemma.

We define:

$$\lambda(x) = \frac{f_1(x)}{f_0(x)}$$

LRT:

$$\psi(X) = \begin{cases} 1, & \text{reject } H_0 \text{ if } \lambda > c \\ 0, & \text{accept } H_0 \text{ if } \lambda \le c \end{cases}$$

Under H_0: $\alpha = [illegible]$

Under H_1: $\beta = [illegible]$

[illegible] constraint [illegible] minimizes $P(H_1)$ [illegible]

Ex. Suppose under hypothesis H_0, observation X has $N(\mu_0, \sigma^2)$, under H_1: $N(\mu_1, \sigma^2)$, where parameters are known.

The likelihood ratio:

$$\lambda(x) = \frac{\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(x-\mu_0)^2}{2\sigma^2}\right\}}$$

$$= \exp\left\{\left(\frac{\mu_1-\mu_0}{\sigma^2}\right)\left(x - \frac{\mu_0+\mu_1}{2}\right)\right\}$$

[illegible] if $X \ge \frac{\mu_0+\mu_1}{2}$, so the test is:

$X > \frac{\mu_0+\mu_1}{2}$ reject H_0

$X \le \frac{\mu_0+\mu_1}{2}$ accept H_0

The LRT for a general threshold τ can be expressed in terms of the log-likelihood:

$$\ln \mathcal{L}(X) \begin{cases} > \ln\tau & \text{, declare } H_1 \\ < \ln\tau & \text{, declare } H_0 \end{cases}.$$

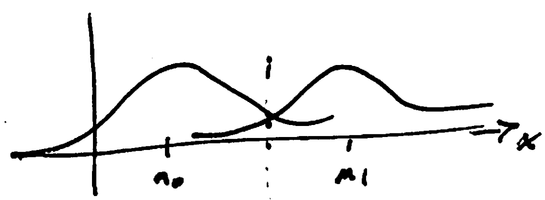
This can be expressed as threshold test for X:

$$X \begin{cases} > \left(\frac{\sigma^2}{m_1-m_0}\right)\ln\tau + \frac{m_0+m_1}{2} & \text{declare } H_1 \\ < \left(\frac{\sigma^2}{m_1-m_0}\right)\ln\tau + \frac{m_0+m_1}{2} & \text{declare } H_0. \end{cases}$$

for MAP, $\tau = \frac{\pi_0}{\pi_1}$, so

$$X \begin{cases} > \gamma_{MAP} & \text{declare } H_1 \\ < \gamma_{MAP} & \text{declare } H_0 \end{cases} \quad \text{where } \gamma_{MAP} = \left(\frac{\sigma^2}{m_1-m_0}\right)\ln\left(\frac{\pi_0}{\pi_1}\right) + \frac{m_0+m_1}{2}.$$

Notice in the Gaussian case, good tests are of the form of comparing X to a threshold



To compute error probabilities, let us think about tests of this form: $X \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$

$$P_F = P(X > \gamma \mid H_0) = P\left(\frac{X-m_0}{\sigma} > \frac{\gamma-m_0}{\sigma} \,\Big|\, H_0\right) = Q\left(\frac{\gamma-m_0}{\sigma}\right)$$

$$P_M = P(X < \gamma \mid H_1) = P\left(\frac{X-m_1}{\sigma} < \frac{\gamma-m_1}{\sigma} \,\Big|\, H_1\right) = Q\left(\frac{m_1-\gamma}{\sigma}\right)$$

$P_e = \pi_0 P_F + \pi_1 P_M$.

The LRT for a general threshold T can be expressed in terms of the log-likelihood:

$$\ln L(X) \begin{cases} > \ln T, & \text{declare } H_1 \\ < \ln T, & \text{declare } H_0 \end{cases}$$

This can be expressed as a threshold test on X:

$$X \begin{cases} > \frac{m_0+m_1}{2} + \ln T \left(\frac{\sigma^2}{m_1-m_0}\right) & \text{declare } H_1 \\ < \frac{m_0+m_1}{2} + \ln T \left(\frac{\sigma^2}{m_1-m_0}\right) & \text{declare } H_0 \end{cases}$$

For MAP, $T = \frac{\pi_0}{\pi_1}$, so

$$X \begin{cases} > x_{MAP} & \text{declare } H_1 \\ < x_{MAP} & \text{declare } H_0 \end{cases}$$

where $x_{MAP} = \frac{m_0+m_1}{2} + \left(\frac{\pi_0}{\pi_1}\right) \ln \left(\frac{\sigma^2}{m_1-m_0}\right)$.

Notice in the Gaussian case, [good to look at the threshold on X] a threshold.

$$X \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

To compute error probabilities, let us think about tests of this form.

$$P_F = P(X > \gamma \mid H_0) = P\left(\frac{X-m_0}{\sigma} > \frac{\gamma-m_0}{\sigma} \,\middle|\, H_0\right) = Q\left(\frac{\gamma-m_0}{\sigma}\right)$$

$$P_M = P(X < \gamma \mid H_1) = P\left(\frac{X-m_1}{\sigma} < \frac{\gamma-m_1}{\sigma} \,\middle|\, H_1\right) = Q\left(\frac{m_1-\gamma}{\sigma}\right)$$

$$P_e = \pi_0 P_F + \pi_1 P_M$$

error probabilities for ML rule are:

$$P_f = P_m = P_e = Q\left(\frac{m_1 - m_0}{2\sigma}\right)$$

function of how separated the two signals are compared to noise std.

Failure Rate functions

related closely to survival function: let T denote lifetime of of a system, where T is positive r.v. with f_T.

$$h(t) \triangleq \lim_{\varepsilon \to 0} \frac{P(t < T \leq t+\varepsilon \mid T > t)}{\varepsilon}$$

Given item working after t time units, the probability the item fails within next ε time units is $h(t)\varepsilon + o(\varepsilon)$.

$$h(t) = \lim_{\varepsilon \to 0} \frac{F_T(t+\varepsilon) - F_T(t)}{(1 - F_T(t))\,\varepsilon}$$

$$= \frac{1}{(1 - F_T(t))}\left(\lim_{\varepsilon \to 0} \frac{F_T(t+\varepsilon) - F_T(t)}{\varepsilon}\right)$$

$$= \frac{f_T(t)}{1 - F_T(t)}$$

Ex: Find failure rate of exponential r.v. with λ.

$$f_T(t) = \lambda e^{-\lambda t} \qquad 1 - F(t) = e^{-\lambda t}$$

so $h(t) = \lambda$ i.e. constant failure rate (memorylessness property).

Transition probabilities for the states :

$$P_d = P_m = P_c = \left(\frac{1}{2}\right)\left(\frac{m_1 - m_2}{2 \cdot s}\right)$$

Fraction of time separated (the two segments are component breaking side).

Failure Rate Functions

Useful identity in survival functions: let T denote lifetime of a system, where T is random var. with F_T.

$$h(t) \triangleq \lim_{\epsilon \to 0} \frac{P(t < T \le t+\epsilon \mid T > t)}{\epsilon}$$

Given that working after t time units, the probability that item fails within next ϵ time units is $h(t)\epsilon + o(\epsilon)$.

$$h(t) = \lim_{\epsilon \to 0} \frac{F_T(t+\epsilon) - F_T(t)}{\epsilon\,(1 - F_T(t))}$$

$$= \frac{1}{1 - F_T(t)} \left(\lim_{\epsilon \to 0} \frac{F_T(t+\epsilon) - F_T(t)}{\epsilon} \right)$$

$$= \frac{f_T(t)}{1 - F_T(t)}$$

Ex: Find failure rate of exponential r.v. with λ.

$$f_T(t) = \lambda e^{-\lambda t} \qquad 1 - F_T(t) = e^{-\lambda t}$$

so $h(t) = \lambda$ i.e. constant failure rate (memoryless property).