$\sigma^2 = Var(X)$

$$Var(X) = E(X^2) - E(X)^2$$

$$Var(\alpha + \beta x) = \beta^2 Var(X)$$

$$Var\left(\frac{1}{n}\sum_{i=1}^{N} X_i\right) = \frac{1}{n} Var(X_1) \rightarrow iid$$

Ropes example: length A & B

$$\hat{a} = \frac{X+Y}{2} \qquad \hat{b} = \frac{X-Y}{2}$$

$E[\hat{a}]$ unbiased

$E[\hat{b}]$ unbiased

$$Var(\hat{a}) = \frac{1}{4} Var(X+Y) = \frac{1}{4} 2\sigma^2 = \frac{1}{2}\sigma^2$$

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

$$Corr = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} \in [-1,1]$$

PDF uniform distribution:

$$F(X) = \int_a^X f(z)\,dz = \frac{X-a}{b-a}$$

$$f(z) = \frac{1}{b-a}$$

random variable: exp. rand.

$$f(x) = \lambda e^{-\lambda x}$$

$$F(x) = 1 - e^{-\lambda x} \quad x \geq 0$$

$$P(a \leq x \leq b) = \int_a^b f(z)\,dz = e^{-\lambda a} - e^{-\lambda b}$$

random variable: memory-less

expectation cont. rand. variable:

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx \xrightarrow{X \sim \mathcal{E}(\lambda)} E[X] = \int_0^{\infty} e^{-\lambda x}\,dx = \frac{1}{\lambda}$$

Joint Distributions

$$F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\,dx\,dy$$

$$f(x,y) = \frac{\delta^2}{\delta x \delta y} F(x,y)$$

Normal:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

Multivariate Normal:

Density fct:

$$\prod_{i=1}^{n} \phi(x_i) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_1^n x_i^2} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}x^T x}$$

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}x_i^2}$$

Bivariate normal (n = 2 multi variate normal)

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \rightarrow \text{Covariance}$$

$$\Sigma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{pmatrix}$$

PDF ($\mu = 0$)

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2(1-\rho)^2}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)}$$

Gamma distribution: model continious variables with skewed distributions

$$f(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$$

MSE: $E_\theta[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + \text{bias}^2(\hat{\theta})$

Sufficiency

Factorization: $f_X(x \mid \theta) = g(T(x), \theta)\, h(\theta)$

if T is sufficient, it doesn't lose info about θ

(minimal sufficient)

↳ if $T(x)$ is minimal if it is a function of every other statistic $\Rightarrow T'(x) = T'(y) / T(x) = T(y)$

-> Very efficient, less memory

Rao Blackwell theorem:

if T is sufficient for θ ∧ $\hat{\theta}$ is an estimator of θ, where for all θ $E(\tilde{\theta}) < \infty$

Let $\tilde{\theta}(x) = E[\hat{\theta}(X) | T(x)] = T(x)$

for all cases of θ

$E[(\hat{\theta} - \theta)] \leq E[(\tilde{\theta} - \theta)]$

Likelihood

$$like(\theta) = \prod_{i=1}^{n} f_x(x_i | \theta)$$

$$loglike(\theta) = \prod_{i=1}^{n} \log f_x(x_i | \theta)$$

when Bernoulli:

$$l(p) = loglike(p) = \left(\sum x_i\right) \log p + \left(n - \sum x_i\right) \log(1-p)$$

$$\frac{dl}{dp} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p}, \text{ when } p = \frac{\sum x_i}{n} = 0$$

and is unbiased

MLE: fct of sufficient statistic

Bayesian estimation

$\pi(\theta)$ prior

$$\pi(\theta|x) = \frac{f_X(x|\theta) \quad \pi(\theta)}{f_X(x)}$$

mostly avoid $f_X(x)$

observe: $\underbrace{\pi(\theta|x)}_{\text{posterior}} \propto \underbrace{f_X(x|\theta)}_{\text{likelihood}} \underbrace{\pi(\theta)}_{\text{prior}}$

next step: estimator (loss function)

posterior loss: (for b)

$$h(b) = \int L(\theta, b)\, \pi(\theta|x)\, d\theta$$

Quadratic loss:

$$h(b) = \int (\theta - b)^2\, \pi(\theta|x)\, d\theta$$

$\rightarrow$ Integration: always $= 1$

LSQ:

$$S(\beta) = \|y - X\beta\|^2$$
$$= (y - X\beta)^T (y - X\beta)$$
$$= \sum_{i=1}^{n} (y_i - x_{i,j}\beta_j)^2$$

to minimize: $\left.\frac{\delta S}{\delta \beta_k}\right|_{\beta=\hat{\beta}} = 0$

So:

$$-2x_{i,k}(y_i - x_{i,j}\hat{\beta}_j) = 0$$
$$x_{i,k}\, x_{i,j}\, \hat{\beta}_j = x_{i,k}\, y_i$$

With matrix form:

$$X^T X \hat{\beta} = X^T Y$$

$X^T X$: positive, semi-definite & has inverse

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$E(\hat{\beta}) = (X^T X)^{-1} X^T E[Y] = (X^T X)^{-1} X^T X \beta = \beta$$

$\Rightarrow$ unbiased