# Probability and Statistics

## Set-Based Probability

### Conditions of Events

- sample space $\Omega$ is an event
- $A$ is an event $\Rightarrow A^c$ is an event
- $A_1, A_2, \ldots$ are events $\Rightarrow \bigcup_{j=1}^{\infty} A_j$ is an event.

### Operations of Set

$$(A \cup B)^c = A^c \cap B^c$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A = (A \cap B) \cup (A \cap B^c)$$

### Definition of Probability

A map from the set of all events to $[0, 1]$

- For every event $A$, $\mathbf{P}(A) \geq 0$
- For the whole sample space $\Omega$, $\mathbf{P}(\Omega) = 1$
- For any infinite sequence of disjoint events $A_1, A_2, \ldots$

$$\mathbf{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbf{P}(A_j)$$

### Simple Sample Space

$$\Omega = \{s_1, \ldots, s_n\},\ \mathbf{P}(\{s_i\}) = \frac{1}{n}$$

### Conditional Probability

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

### Law of Total Probability

$B_1, \ldots, B_k$ form a partition of $\Omega$

$$\mathbf{P}(A) = \sum_{j=1}^{k} \mathbf{P}(B_j)\,\mathbf{P}(A \mid B_j)$$

### Independent Events

Two events $A$ and $B$ are independent if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$$

$A_1, \ldots, A_k$ are mutually independent if

$$\text{For any } \mathcal{M} \in \{1, \ldots, k\}, \mathbf{P}\left(\bigcap_{j \in \mathcal{M}} A_j\right) = \prod_{j \in \mathcal{M}} \mathbf{P}(A_j)$$

### Bayes' Theorem

$B_1, \ldots, B_k$ form a partition of $\Omega$

$$\mathbf{P}(B_i \mid A) = \frac{\mathbf{P}(B_i)\,\mathbf{P}(A \mid B_i)}{\sum_{j=1}^{k} \mathbf{P}(B_j)\,\mathbf{P}(A \mid B_j)}$$

## Random Variables and Distribution

### Random Variable

Assign a number to each outcome of a random experiment

> Consider a random experiment with a sample space $\Omega$. A function $X$, which assigns to each element $s \in \Omega$ one and only one number $X(s) = x$, is called a **random variable**. The **space** or **range** of $X$ is the set of real numbers $\mathcal{D} = \{x : x = X(s), s \in \Omega\}$.

### Discrete Distribution

Probability Mass Function (pmf)

$$p_X(d) = \mathbf{P}(\{s \in \Omega : X(x) = d\})$$

Bernoulli Distribution with parameter $p$

$$p_X(1) = p \text{ and } p_X(0) = 1 - p$$

Uniform Distribution on Integers $[a, b] \cap \mathbb{Z}$

$$p_X(x) = \frac{1}{b - a + 1} \qquad \text{for } x = a, \ldots, b$$

Binomial Distribution with parameter $n$ and $p$

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{for } x = 0, 1, \ldots, n$$

Geometric Distribution with parameter $p$

$$p_X(x) = (1-p)^{x-1} p \qquad \text{for } x = 1, 2, \ldots$$

### Continuous Distribution

Probability Density Function (pdf)

$$\mathbf{P}(a \leq X \leq b) = \int_a^b f(x)dx$$

Cumulative Distribution Function (cdf)

$$F_X(x) = \mathbf{P}(X \leq x) \qquad \text{for } -\infty < x < \infty$$
$$F_X(x) = \sum_{d_i \leq x} p_X(d_i) \text{ or } \int_{-\infty}^{x} f(x)dx$$

### Joint Distribution

joint pdf (similarly joint pmf and mixed joint distribution)

$$\mathbf{P}(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{X,Y}(x,y)dxdy$$

joint cdf

$$F_{X,Y}(x,y) = \mathbf{P}(X \leq x \text{ and } Y \leq y)$$

marginal cdf of $X$

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y)$$

marginal pdf(pmf) of $X$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

## Independent Random Variables

Two random variables $X$ and $Y$ are independent iff

$$F_{X,Y}(x,y) = F_X(x)F_y(y)$$
or
$$f_{X,Y}(x,y) = f_X(x)f_y(y)$$

## Conditional Distribution

conditional distribution of $X$ given $Y = y$

$$g_X(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

## Independent and Identically Distributed

A collection of random variables $(X_i)_{i=1}^n$ is **i.i.d.**

- each $X_i$ has the same probability distribution
- they are mutually independent

## Function of Variables

$Y = r(X)$, $r$ is differentiable and one-to-one

$$f_Y(y) = f_X(r^{-1}(y)) \left| \frac{d}{dy} r^{-1}(y) \right|$$

$Y = a_1X_1 + a_2X_2 + b$, $a_1 \neq 0$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1,X_2}\left(\frac{y - b - a_2x_2}{a_1}, x_1\right) \frac{1}{|a_1|} dx_2$$

$Y_j = r_j(\boldsymbol{X})$, inverse transformation $x_i = s_i(\boldsymbol{y})$

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{s})|J|$$

$$J = \det \begin{bmatrix} \frac{\partial s_1}{\partial y_1} & \cdots & \frac{\partial s_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_n}{\partial y_1} & \cdots & \frac{\partial s_n}{\partial y_n} \end{bmatrix}$$

## Random Process

A sequence of random variables $X_1, X_2, \dots$ is called a stochastic process or random process with discrete time parameter. $X_1$ is called the initial state.

## Markov Chain

the conditional distribution of $X_{n+1}$ depend only on $X_n$

$$\mathbf{P}(X_{n+1} \leq b \mid X_i = x_i) = \mathbf{P}(X_{n+1} \leq b \mid X_n = x_n)$$

Stochastic Matrix is a square matrix whose entries are all nonnegative and the sum of each row is 1.
Transition Matrix is a $k \times k$ martix($k$ is the number of possible states), $p_{ij} = \mathbf{P}(X_{n+1} = j \mid X_n = i)$

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{bmatrix}$$

state vector provided initial observation $\boldsymbol{x}^{(0)}$

$$\boldsymbol{x}^{(n)} = \begin{bmatrix} \mathbf{P}(X_n = 1) & \cdots & \mathbf{P}(X_n = k) \end{bmatrix} = \boldsymbol{x}^{(0)} P^n$$

# Numerical Characteristics

## Expectation

$$\mathbf{E}(X) = \sum x p_X(x) \text{ or } \int_{-\infty}^{\infty} x f(x) dx$$

expectation exists if one of the two integrals converges

$$\int_{-\infty}^{0} x f(x) dx \quad \text{and} \quad \int_{0}^{\infty} x f(x) dx$$

$$\mathbf{E}(r(X)) = \int_{-\infty}^{\infty} r(x) f_X(x) dx$$

$$\mathbf{E}(aX + b) = a\mathbf{E}(X) + b$$

$$\mathbf{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbf{E}(X_i)$$

if $X_1, \dots, X_n$ are independent and $\mathbf{E}(X_i)$ is finite

$$\mathbf{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbf{E}(X_i)$$

## Jensen's Inequality

Let $f$ be a convex function

$$\mathbf{E}(f(X)) \geq f(\mathbf{E}(X))$$

## Variance

$$\mathbf{Var}(X) = \mathbf{E}\left((X - \mu)^2\right) = \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2$$

standard deviation $\sigma = \sqrt{\mathbf{Var}(X)}$

$$\mathbf{Var}(aX + b) = a^2 \mathbf{Var}(X)$$

if $(X_i)_{i=1}^n$ are independent variables with finite means

$$\mathbf{Var}\left(\sum_{i=1}^n a_i X_i + b\right) = \sum_{i=1}^n a_i^2 \mathbf{Var}(X_i)$$

## Moment and MGF

$k$th (raw) moment $\quad \mathbf{E}\left(X^k\right)$
$k$th central moment $\quad \mathbf{E}\left((X - \mu)^k\right)$
$k$th moment exists iff $\mathbf{E}\left(|X|^k\right) < \infty$
moment generating function(mgf)

$$M_X(t) = \mathbf{E}\left(e^{tX}\right)$$

$$M_X(0)^{(k)} = \mathbf{E}\left(X^k\right)$$

if random variables have the same mgf, then they must follow the same distribution.
$Y = X_1 + \cdots + X_n$ and $(X_i)_{i=1}^n$ are independent

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$$

## Skewness

the lack of symmetry $\quad \mathbf{E}\left((X - \mu)^3 / \sigma^3\right)$

## Minimizing the Error

mean $\mu = \mathbf{E}(X)$
median $\mathbf{P}(X < m) \geq \frac{1}{2}$ and $\mathbf{P}(X \geq m) \geq \frac{1}{2}$
Mean Squared Error(MSE) $\quad \mathbf{E}\left((X - d)^2\right) \geq \mathbf{E}\left((X - \mu)^2\right)$
Mean Absolute Error(MAE) $\quad \mathbf{E}(|X - d|) \geq \mathbf{E}(|X - m|)$

## Covariance and Correlation

$\mathrm{Cov}(X,Y) = \mathbf{E}((X-\mu_X)(Y-\mu_Y)) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$

correlation of $X, Y$ describes how they are linearly related

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

$$\mathbf{Var}\left(\sum_{i=1}^n a_i X_i + b\right) = \sum_{i=1}^n a_i^2 \mathbf{Var}(X_i) + \sum_{i<j} 2a_i a_j \mathrm{Cov}(X_i, X_j)$$

Cauchy-Schwarz Inequality

$$\mathbf{E}(XY)^2 \le \mathbf{E}\left(X^2\right)\mathbf{E}\left(Y^2\right)$$

$$\mathrm{Cov}(X,Y)^2 \le \sigma_X^2 \sigma_Y^2$$

## Conditional Expectation

$$\mathbf{E}(Y \mid X = x) = \int_{-\infty}^{\infty} y f_Y(y \mid x) dy$$

$$\mathbf{E}(Y) = \mathbf{E}(\mathbf{E}(Y \mid X))$$

$$\mathbf{Var}(Y \mid X) = \mathbf{E}\left((Y - \mathbf{E}(Y \mid X))^2 \mid X\right) = \mathbf{E}\left(Y^2 \mid X\right) - \mathbf{E}(Y \mid X)^2$$

$$\mathbf{Var}(Y) = \mathbf{E}(\mathbf{Var}(Y \mid X)) + \mathbf{Var}(\mathbf{E}(Y \mid X))$$

$$d(X) = \mathbf{E}(Y \mid X) \text{ minimizes } \mathbf{E}\left((Y - d(X))^2\right)$$

## Some Math

Gaussian Intergral

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

Gamma Function

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \qquad \Gamma(n) = (n-1)!$$

Beta Function

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Cauchy Distribution

$$\mathrm{Cauchy}(1,0) \sim t(1) \quad f(x) = \frac{1}{\pi(1+x^2)}$$

$$f(x) = \frac{1}{\pi}\frac{\gamma}{(x-\mu)^2+\gamma^2} \qquad \text{for } -\infty < x < \infty$$

$$\sum_{i=1}^n \mathrm{Cauchy}(\gamma_i, \mu_i) \sim \mathrm{Cauchy}\left(\sum_{i=1}^n \gamma_i, \sum_{i=1}^n \mu_i\right)$$

# Special Distributions

## Bernoulli Distribution

$$p_X(x) = p^x(1-p)^{1-x} \qquad \text{for } x = 0, 1$$

$$\mathbf{E}(X) = p \qquad \mathbf{Var}(X) = p(1-p)$$

$$M_X(t) = 1 - p + pe^t \qquad \text{for } t \in \mathbb{R}$$

## Binomial Distribution

Number of success in $n$ trials $\mathrm{Ber}(p)$

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{for } x = 0, 1, \ldots, n$$

$$\mathbf{E}(X) = np \qquad \mathbf{Var}(X) = np(1-p)$$

$$M_X(t) = (1 - p + pe^t)^n \qquad \text{for } t \in \mathbb{R}$$

## Hypergeometric Distribution

Number of colored items drawing $n$ in $N$ with $M$ colored

$$\lim_{n\to\infty} \frac{M_n}{N_n} = p \longrightarrow \lim_{n\to\infty} \mathrm{Hyp}(N_n, M_n, n) \sim \mathrm{Bin}(n, p)$$

$$p_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \text{for } \max(0, n-N+M) \le x \le \min(n, M)$$

$$\mathbf{E}(X) = \frac{nM}{N} \qquad \mathbf{Var}(X) = \frac{nM}{N}\frac{(N-M)(N-n)}{N(N-1)}$$

## Poisson Distribution

Number of events occurring in unit time with average $\lambda$

$$\lim_{n\to\infty} np_n = \lambda \longrightarrow \lim_{n\to\infty} \mathrm{Bin}(n, p_n) \sim \mathrm{Poi}(\lambda)$$

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{for } x = 0, 1, \ldots$$

$$\mathbf{E}(X) = \lambda \qquad \mathbf{Var}(X) = \lambda$$

$$M_X(t) = e^{\lambda(e^t - 1)} \qquad \text{for } t \in \mathbb{R}$$

## Negative Binomial Distribution

Number of failures to get the $r$th success in $\mathrm{Ber}(p)$

$$p_X(x) = \binom{r+x-1}{x} p^r (1-p)^x \qquad \text{for } x = 0, 1, \ldots$$

$$\mathbf{E}(X) = \frac{r(1-p)}{p} \qquad \mathbf{Var}(X) = \frac{r(1-p)}{p^2}$$

$$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r \qquad \text{for } t < \ln\frac{1}{1-p}$$

## Geometric Distribution

$\mathrm{Geo}(p) \sim \mathrm{Negbin}(1, p)$. Memoryless Property

$$\mathbf{P}(X = k + t \mid X \ge k) = \mathbf{P}(X = t) \qquad (k, t \ge 0)$$

## Normal Distribution

$$X_i \sim N(\mu_i, \sigma_i^2) \longrightarrow \beta + \sum_{i=1}^n \alpha_i X_i \sim N\left(\beta + \sum_{i=1}^n \mu_i, \sum_{i=1}^n \alpha_i \sigma_i^2\right)$$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{for } x \in \mathbb{R}$$

$$\mathbf{E}(X) = \mu \qquad \mathbf{Var}(X) = \sigma^2$$

$$M_X(t) = e^{\frac{1}{2}\sigma^2 t^2 + \mu t} \qquad \text{for } t \in \mathbb{R}$$

Standard Normal Distribution $\phi(x) \sim \mathrm{N}(0,1)$
Lognormal Distribution $\ln(X) \sim \mathrm{N}(\mu, \sigma^2)$

## Gamma Distribution

Time elasped until $\alpha$th occurrence of event in $\text{Poi}(\beta)$

$$k\Gamma(\alpha,\beta) \sim \Gamma\left(\alpha, \frac{\beta}{k}\right)$$

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad \text{for } x > 0$$

$$\mathbf{E}(X) = \frac{\alpha}{\beta} \qquad \mathbf{Var}(X) = \frac{\alpha}{\beta^2}$$

$$M_X(t) = \left(\frac{\beta}{\beta - t}\right)^\alpha \qquad \text{for } t < \beta$$

## Exponential Distribution

$\text{Exp}(\beta) \sim \Gamma(1,\beta)$. Memoryless Property

$$\mathbf{P}(X \geq t + h \mid X \geq h) = \mathbf{P}(X \geq t) \qquad (h, t \geq 0)$$

# Large Random Samples

## Inequalities

Markov Inequality $X$ is nonnegative, $t > 0$

$$\mathbf{P}(X \geq t) \leq \frac{\mathbf{E}(X)}{t}$$

Chebyshev Inequality $t > 0$

$$\mathbf{P}(|X - \mu| \geq t) \leq \frac{\mathbf{Var}(X)}{t^2}$$

Chernoff Bound for any $a \in \mathbb{R}$

$$\mathbf{P}(X \geq a) \leq \min_{t>0} \frac{M_X(t)}{e^{at}} \left(\leq \frac{\mathbf{E}\left(e^{tX}\right)}{e^{at}}\right)$$

## Convergence

$$X_n \xrightarrow{P} X \quad \lim_{n\to\infty} \mathbf{P}(|X_n - X| < \varepsilon) = 1$$

$$X_n \xrightarrow{d} X \quad \lim_{n\to\infty} F_n(x) = F(x)$$

$$X_n \xrightarrow{a.s.} X \quad \mathbf{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Almost Sure $\Rightarrow$ in Distribution $\Rightarrow$ in Probability

## Law of Large Numbers(LLN)

$(X)_{i=1}^n$ (i.i.d.) with mean $\mu$ and finite variance

$$\overline{X}_n \xrightarrow{P} \mu \text{ (Weak) or } \overline{X}_n \xrightarrow{a.s.} \mu \text{ (Strong)}$$

## Central Limit Theorem(CLT)

random sample $(X)_{i=1}^n$ with mean $\mu$ and variance $\sigma^2$

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$$

## Delta Method

$T$ is a statistic of random sample $(X)_{i=1}^n$

$$a_n(T - \theta) \xrightarrow{d} F \Longrightarrow a_n \frac{g(T) - g(\theta)}{g'(\theta)} \xrightarrow{d} F$$

Take $T = \overline{X}_n$, $\theta = \mu$, $a_n = \frac{\sqrt{n}}{\sigma}$, $F = \Phi$ and $g(x) = \frac{1}{x}$

$$\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) \xrightarrow{d} \Phi \Longrightarrow -\frac{\sqrt{n}\mu^2}{\sigma}\left(\frac{1}{\overline{X}_n} - \frac{1}{\mu}\right) \xrightarrow{d} \Phi$$

# Estimation

## Method of Moments

$k$ parameters $\boldsymbol{\theta}$ and first $k$ moments $\boldsymbol{\mu}$

$$\boldsymbol{\mu} = \left(\mathbf{E}(X), \ldots, \mathbf{E}\left(X^k\right)\right) = g(\boldsymbol{\theta}) \Longrightarrow \boldsymbol{\theta} = g^{-1}(\boldsymbol{\mu})$$

Then use sample moments $\hat{\boldsymbol{\mu}}$ to estimate $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}} = g^{-1}(\hat{\boldsymbol{\mu}}) = g^{-1}\left(\left(\sum_{i=1}^k x_i, \ldots, \sum_{i=1}^k x_i^k\right)\right)$$

## Maximum Likelihood Estimation

Likelihood Function

$$L(\theta) = \prod_{i=1}^n f(x_i;\theta) \left(= \prod_{i=1}^n f(x_i \mid \theta)\right)$$

Maximum Likelihood Estimator(MLE)

$$\hat{\theta} = \arg\max L(\theta)$$

$$\hat{\theta} \text{ is MLE of } \theta \Longrightarrow g(\hat{\theta}) \text{ is MLE of } g(\theta)$$

## Unbiased and Consistent

Unbiasedness $\mathbf{E}\left(\hat{\theta}\right) = \theta$

Consistency $\hat{\theta} \xrightarrow{P} \theta$

$$\lim_{n\to\infty} \mathbf{E}\left(\hat{\theta}\right) = \theta, \lim_{n\to\infty} \mathbf{Var}\left(\hat{\theta}\right) = 0 \Longrightarrow \text{consistent}$$

sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ are biased and consistent

## Bayesian Statistics

Prior Distribution $h(\theta)$ Posterior Distribution $h(\theta \mid \boldsymbol{x})$

$$h(\theta \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x};\theta)}{g(\boldsymbol{x})} \left(= \frac{f(\boldsymbol{x},\theta)}{g(\boldsymbol{x})}\right) = \frac{h(\theta)\prod_{i=1}^n f(x_i \mid \theta)}{\int_\Omega h(\theta)\prod_{i=1}^n f(x_i \mid \theta) d\theta}$$

Conjugate Family

$$\lambda \sim \Gamma(\alpha,\beta), X|\lambda \sim \text{Poi}(\lambda) \Rightarrow \lambda|X \sim \Gamma(\alpha + \sum_{i=1}^n x_i, \beta + n)$$

## Bayes Estimators

Loss Function $L(\theta, \hat{\theta})$

Bayes Estimator $\hat{\theta}(\boldsymbol{X})$ minimizes $\mathbf{E}\left(L(\theta,\hat{\theta}) \mid \boldsymbol{X}\right)$

squared error loss $L(\theta,\hat{\theta}) = (\theta - \hat{\theta})^2 \Longrightarrow \hat{\theta} = \mathbf{E}(\theta \mid \boldsymbol{X})$

# Distributions of Estimators

## Chi-Square Distribution

$$\chi^2(m) \sim \Gamma(\frac{m}{2}, \frac{1}{2})$$

$$X \sim N(0,1) \Longrightarrow X^2 \sim \chi^2(1)$$

## t Distribution

$$Z \sim N(0,1), Y \sim \chi^2(m) \Longrightarrow \frac{Z}{\sqrt{Y/m}} \sim t(m)$$

## F Distribution

$$Y \sim \chi^2(m), W \sim \chi^2(n) \Longrightarrow \frac{Y/m}{W/n} \sim F(m,n)$$

## Confidence Interval

Confidence Interval with coefficient $1-\alpha$ $(L, U)$

$$\mathbf{P}\left(L < g(\theta) < U\right) \geq 1-\alpha$$

Pivotal Quantity $Z = g(\boldsymbol{X}, \theta)$ with distribution independent of $\theta$

$$\text{CI of } Z, \theta = r(Z) \Longrightarrow \text{CI of } \theta$$

## Pivot of Normal distribution

(Only) For $N(\mu, \sigma^2)$, $\overline{X}_n$ and $S_n^2$ are independent
Estimate $\mu$ with known $\sigma^2$

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Estimate $\mu$ with unknown $\sigma^2$

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$$

Estimate $\sigma^2$ with known $\mu$

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$$

Estimate $\sigma^2$ with unknown $\mu$

$$(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

## Order Statistics

$Y_i = X_{(i)}$ is the $i$th smallest of random sample $(X)_{i=1}^n$
$f$ and $F$ are the pdf and cdf of $X$

$$g(y_1, \dots, y_n) = n! f(y_1) \dots f(y_n) \qquad \text{for } y_1 < \dots < y_n$$

$$g(y_i) = n\binom{n-1}{k-1} F(y_i)^{i-1}(1-F(y_i))^{n-i} f(y_i)$$

$$g(y_i, y_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F(y_i)^{i-1}$$
$$(F(y_j)-F(y_i))^{j-i-1}(1-F(y_j))^{n-j} f(y_i) f(y_j) \text{ for } y_i < y_j$$

# Hypothesis Testing

## Hypotheses

Null Hypothesis $H_0$ and Alternative Hypothesis $H_1$

$$H_0 : \theta \in \omega_0 \quad H_1 : \theta \in \omega_1 \quad (\text{partition of } \Omega)$$

Type I Error $\theta \in \omega_0$ but reject $H_0$
Type II Error $\theta \in \omega_1$ but retain(cannot reject) $H_0$

## Controlling Type I Error

Critical Region reject $H_0$ if $\boldsymbol{X} \in C$
Rejection Region reject $H_0$ if $T = r(\boldsymbol{X}) \in R$
Power Function The probability that $\theta$ will reject $H_0$

$$\pi(\theta) = \mathbf{P}\left(\boldsymbol{X} \in C \mid \theta\right)$$

Significance Level of test or size of $C$ is $\alpha$

$$\text{for } \theta \in \omega_0, \quad \pi(\theta) \leq \sup_{\theta \in \omega_0} \mathbf{P}\left(\boldsymbol{X} \in C \mid \theta\right) = \alpha$$

## Testing by p-value

p-value is the probability of obtaining results at least as extreme as $\boldsymbol{X}$ assuming $H_0$ is correct
if rejection region is $(c, +\infty)$ then for observation $t = T(\boldsymbol{x})$

$$p = \sup_{\theta \in \omega_0} \mathbf{P}\left(T > t \mid \theta\right)$$

if we want to give a level $\alpha$ test, reject $H_0$ if $\alpha \geq p$

## Testing Simple Hypotheses

Simple Hypothesis $\omega_0, \omega_1$ contains only a single value of $\theta$

$$H_0 : \theta = \theta_0 \quad H_1 : \theta = \theta_1$$

Best Critical Region $C$ makes probablilty of Type I Error $\alpha$ and minimizes probablilty of Type II Error

$$\mathbf{P}\left(\boldsymbol{X} \in C \mid H_0\right) = \alpha, \mathbf{P}\left(\boldsymbol{X} \in C \mid H_1\right) \text{ is maximal}$$

Neyman-Pearson $C$(size $\alpha$) is a best critical region if

$$C = \left\{X : \frac{L(\theta_0; \boldsymbol{x})}{L(\theta_1; \boldsymbol{x})} \leq k\right\}$$

## The $\chi^2$ Test

$n$ trials with $k$ possible outcomes

$$H_0 : p_i = p_i^0 \quad H_1 : \text{not } H_0$$

$$Q = \sum_{i=1}^{k} \frac{(N_i - np_i^0)^2}{np_i^0} \xrightarrow{d} \chi^2(k-1)$$