$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

| Distribution | pdf/pmf | cdf | $\mu, \sigma^2$ | cf $\varphi(t)$ | MLE | |
|---|---|---|---|---|---|---|
| $\mathcal{N}(\mu,\sigma^2)$ on $\mathbb{R}$ | $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | $\Phi(\frac{x-\mu}{\sigma})$ | $\mu, \sigma^2$ | $e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$ | $\hat\mu = \bar X_n$, $\hat\sigma^2 = \frac{1}{n}\sum(X_i - \bar X_n)^2$ | CLT, test stats |
| $Unif(a,b)$ on $[a,b]$ | $\frac{1}{b-a}$ | $\frac{x-a}{b-a}$ | $\frac{a+b}{2}, \frac{(b-a)^2}{12}$ | $\frac{e^{itb}-e^{ita}}{it(b-a)}$ | $\hat\theta = \max(X_i)$ $(Unif(0,\theta))$ | Equal probs in interval |
| $Exp(\lambda)$ on $\mathbb{R}_{>0}$ | $\lambda e^{-\lambda x}$ | $1-e^{-\lambda x}$ | $\frac{1}{\lambda}, \frac{1}{\lambda^2}$ | $\frac{\lambda}{\lambda - it}$ | $\hat\lambda = \frac{1}{\bar X_n}$ | Waiting time |
| $Bin(n,p)$ on $0,\dots,n$ | $\binom{n}{k} p^k (1-p)^{n-k}$ | $\sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$ | $np, np(1-p)$ | $(1-p+pe^{it})^n$ | $\hat p = \frac{1}{n}\sum X_i$ $(\frac{\bar X_n}{n})$ | # of successes |
| $Poi(\lambda)$ on $\mathbb{N}_0$ | $\frac{e^{-\lambda}\lambda^k}{k!}$ | $\sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^i}{i!}$ | $\lambda, \lambda$ | $e^{\lambda(e^{it}-1)}$ | $\hat\lambda = \bar X_n$ | Rare events |
| $Ber(p)$ on $\{0,1\}$ | $p^x(1-p)^{1-x}$ | $\frac{x-a}{b-a}$ if $0 \le x \le 1$ | $p, p(1-p)$ | $(1-p)+pe^{it}$ | $\hat p = \bar X_n$ | $Bin(n=1)$ |
| $Geo(p)$ on $\mathbb{N}_{>0}$ | $(1-p)^{k-1}p$ | $1-(1-p)^k$ | $\frac{1}{p}, \frac{1-p}{p^2}$ | $\frac{pe^{it}}{1-(1-p)e^{it}}$ | $\hat p = \frac{1}{\bar X_n}$ | Trials until success |
| $\Gamma(\alpha,\beta)$ on $\mathbb{R}_{>0}$ | $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ | — | $\frac{\alpha}{\beta}, \frac{\alpha}{\beta^2}$ | $(1-it/\beta)^{-\alpha}$ | — | Sum of exponentials |

cdf = pdf $\frac{dx}{d}$
pdf = cdf $\frac{d}{dx}$

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

**De Morgan**

$(\bigcup_{i\in I} A_i)^c = \bigcap_{i\in I} A_i^c$

$(\bigcap_{i\in I} A_i)^c = \bigcup_{i\in I} A_i^c$

**σ-algebras**

1 $\emptyset \in \mathcal{A}$ $\Rightarrow$ if $(A_i)_{i\in\mathbb{N}} \in \mathcal{A}$, then $\bigcap_{i\in\mathbb{N}} A_i \in \mathcal{A}$

2 if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$ $\Rightarrow$ if $A, B \in \mathcal{A}$, then $A\cup B \in \mathcal{A}$, $A\cap B \in \mathcal{A}$

3 if $(A_i)_{i\in\mathbb{N}} \in \mathcal{A}$, then $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$ — $A\setminus B \in \mathcal{A}$, $A\Delta B = (A\setminus B)\cup(B\setminus A) \in \mathcal{A}$

$\Rightarrow$ if $\mathcal{A}_\alpha$ σ-algebras over $\Omega$, then $\bigcap_\alpha \mathcal{A}_\alpha$ is also σ-algebra over $\Omega$

$\mathcal{B}: \sigma(\{(a,b) \subset \mathbb{R} \mid a<b\})$
includes all: closed intervals, half-open intervals, singletons $\{x\}$, countable sets

**measurable space** $(\Omega, \mathcal{A})$ $\mathcal{A}$ σ-a on $\Omega$

**measure space** $(\Omega,\mathcal{A},\mu)$ measurable space + $\mu: \mathcal{A} \to [0,\infty]$ is a measure
when 1 $\mu(\emptyset) = 0$ 2 $\mu(A) \ge 0$ $\forall A \in \mathcal{A}$
3 $\mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mu(A_i)$

**Probability measure** $\mu: \mathcal{A} \to [0,\infty]$
1, 2, 3, 4: $\mu(\Omega) = 1$

**probability space $(\Omega, \mathcal{A}, P)$**
- $\Omega$ sample space
- $\mathcal{A}$ σ-algebra over $\Omega$ (events) (usually $\mathcal{P}(\Omega)$)
- $P: \mathcal{A} \to [0,1]$ with 1 $P(\Omega) = 1$ 2 $P(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} P(A_i)$

**Bayes' theorem**

$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$

**Inclusion-exclusion formula**

$P(A_1 \cup \dots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\le i_1 < \dots < i_k \le n} P(A_{i_1} \cap \dots \cap A_{i_k})$

**Expectation** $\mathbb{E}[X] = \sum_{x\in\Omega} x\, p_X(x)$ / $\mathbb{E}[X] = \int_{\mathbb{R}} x\, p_X(x)\, dx$

$\mathbb{E}[aX+bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$, $\mathbb{E}[aX] = a\mathbb{E}[X]$ $\mathbb{E}[c] = c$

If $X(\omega) \ge 0$ $\forall \omega\in\Omega$, then $\mathbb{E}[X] \ge 0$ $\mathbb{E}[X] = \mathbb{E}[X^2]$ if $X \sim Ber$

If $X(\omega) \ge Y(\omega)$ $\forall\omega\in\Omega$, then $\mathbb{E}[X] \ge \mathbb{E}[Y]$

$\mathbb{E}[XY] = \sum_x \sum_y xy \cdot P(X=x, Y=y)$ or $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ if $X \perp\!\!\!\perp Y$

$\mathbb{E}[X|Y] = \int_{\mathbb{R}} x \frac{p_{X,Y}(x,y)}{p_Y(y)} dx$

$\mathbb{E}[X] = \sum_{n=0}^{\infty} P(X \ge n)$ / $\mathbb{E}[X] = \int_0^\infty 1 - F_X(x)\, dx$

$\mathbb{E}[g(X)] = \sum_{k=0}^{\infty} g(k) P(X=k)$ / $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) p_X(x)\, dx$

**RVRV** is a function $\Omega \to \mathbb{R}$

**Variance** $var[X] \ge 0$

$var[X] = \mathbb{E}[X - \mathbb{E}[X]^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

$var[aX+b] = a^2\, var[X]$

$var[X+Y] = var[X] + var[Y] + \underbrace{2Cov(X,Y)}_{=0 \text{ if } X \perp\!\!\!\perp Y}$

$var[X] = 0 \Leftrightarrow P(X = \mathbb{E}[X]) = 1$

**conditional pdf** $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$

**median** $F_X(x_{0.5}) = 0.5$

**mode** maximum of pdf

**Covariance**

$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

$Cov(X,X) = var[X]$

$Cov(aX+b, Y) = a\cdot Cov(X,Y)$

if $X \perp\!\!\!\perp Y$, then $Cov(X,Y) = 0$

$X, Y$ uncorrelated if $Cov(X,Y) = 0$

if $X \perp\!\!\!\perp Y$, then
- $p_{X,Y}(x,y) = p_X(x)p_Y(y)$
- $Z := X+Y$, $p_Z(z) = \int_{\mathbb{R}} p_X(z-y) p_Y(y)\, dy$

**correlation coefficient** $\rho[X,Y] = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$

if $X \perp\!\!\!\perp Y$, then $\rho[X,Y] = 0$

$\rho = 1$: perfect positive linear relationship
$\rho = 0$: uncorrelated
$\rho = -1$: perfect negative linear relationship

**$X \perp\!\!\!\perp Y$ if:** $p_{X,Y}(x,y) = p_X(x)\, p_Y(y)$ $\forall (x,y) \in \mathbb{R}^2$

formula for affine transformations: $Y = aX+b$ $\Rightarrow$ $p_Y(y) = \frac{1}{|a|} p_X(\frac{y-b}{a})$

**Chebychev's inequality**
$P(|X-\mu| \ge \varepsilon) \le \frac{\sigma^2}{\varepsilon^2}$

**Markov's inequality**
$P(X \ge \varepsilon) \le \frac{\mathbb{E}[X]}{\varepsilon}$

**Jensen's inequality**
$\mathbb{E}[g(X)] \ge g(\mathbb{E}[X])$ if $g$ is strictly convex (eg. $x^2$, $e^x$, $-\ln(x)$, $x^p$ $p>1$)
$\mathbb{E}[g(X)] \le g(\mathbb{E}[X])$ if $g$ is strictly concave (eg. $\ln(x)$, $\sqrt{x}$, $x^p$ $0<p<1$)

## modes of convergence

- almost surely: $(X_n)_{n\in\mathbb{N}}$ conveges to another RVRV on $(\Omega,\mathcal{A},P)$, if $P(\{\omega\in\Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$
- in probability: $(X_n)_{n\in\mathbb{N}}$ converges to another RVRV on $(\Omega,\mathcal{A},P)$, if $\forall \varepsilon > 0: P(|X_n - X| > \varepsilon) \to 0$ as $n\to\infty$
- in distribution: $(X_n)_{n\in\mathbb{N}}$ converges to another RVRV, if $F_{X_n} \to F_X(x)$ for all continuity points of $F_X$

**CLT** If $X_1, X_2, \dots, X_n$ are i.id. RVRVs with mean $\mu$ and variance $\sigma^2 < \infty$, then the normalized sample mean converges in distribution to the standard normal distribution. $Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$ as $n\to\infty$

$$P(X \le a) \approx P\left(Z \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right) \quad \left(\text{with } Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)\right)$$

**WLLN** If " with finite mean $\mu = \mathbb{E}[X_i]$, then the sample mean converges in probability to the true mean. $\bar{X}_n = \frac{1}{n}\sum X_i$ ~~...~~ $\xrightarrow{P} \mu$ as $n\to\infty$.

**SLLN** If ", then the sample mean almost surely converges to the true mean. $\bar{X}_n = \frac{1}{n}\sum X_i \xrightarrow{a.s.} \mu$ as $n\to\infty$

interpretation: · WLLN guarantees that $\bar{X}_n$ gets closer to $\mu$ in probability as $n$ increases.
- SLLN guarantees that $\bar{X}_n$ will actually converge to the true mean $\mu$ for almost every outcome.

## characteristic functions $\varphi_X(t) = \mathbb{E}[e^{itX}]$

$\varphi_X(t) = \varphi_Y(t)\ \forall t\in\mathbb{R} \Rightarrow X =_d Y$ (uniqueness)

$|\varphi_X(t)| \le 1$, $\varphi_X(0) = 1$, $X =_d -X \Leftrightarrow \varphi_X(t) \in \mathbb{R}$

if $X \perp\!\!\!\perp Y$, then $\varphi_{X+Y}(t) = \varphi_X(t) + \varphi_Y(t)$

if $X_i$ i.i.d. RVs, $S_n = \sum X_i$, then $\varphi_{S_n}(t) = \prod_{i=1}^n \varphi_{X_i}(t)$

### Lévy's continuity theorem

if $\varphi_{X_n}(t) \to \varphi_X(t)$ pointwise and $\varphi_X$ continuous at $t=0$, then $X_n \xrightarrow{d} X$

### mean squared error (mse)

$$\text{mse}[\hat{\theta}] = \mathbb{E}[(\theta - \hat{\theta})^2] = \text{Var}[\hat{\theta}] + \text{Bias}[\theta]^2$$

## statistical model $(\Omega, \mathcal{P}(\Omega), P_\Theta)$

**likelihood function** $L(\theta \mid x_1,\dots,x_n) = \prod_{i=1}^n \underbrace{p}_{\text{just pdf}}(x_i \mid \underbrace{\theta}_{\text{unknown parameter}})$

**log likelihood function** $\ell(\theta \mid x_1,\dots,x_n) = \sum_{i=1}^n \ln(p(x_i \mid \theta))$

**maximum likelihood estimator (mle)**

$\theta_{mle} = \text{argmax}_{\theta\in\Theta} L(\theta \mid x_1,\dots,x_n)$ / $_{\theta\in\Theta}\,\ell(\theta \mid x_1,\dots,x_n)$

$\Rightarrow$ compute $\frac{d\ell}{d\theta}$ and set to zero, confirm $\frac{d^2\ell}{d\theta^2} < 0$

- biased estimator if $\mathbb{E}[\hat{\theta}] \neq \theta$
- unbiased estimator if $\mathbb{E}[\hat{\theta}] = \theta \quad \forall \theta \in \Theta$

$\text{Bias}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta$

$z_\alpha = P(Z \ge z_\alpha) = \alpha \ \hat{=}\ P(Z \le z_\alpha) = 1-\alpha$ $\quad z_{0,05} \approx 1{,}645$, $z_{0,025} \approx 1{,}96$

| | fail to reject $H_0$ | reject $H_0$ |
|---|---|---|
| $H_0$ is true | correct decision | type I error |
| $H_1$ is true | type II error | correct decision |

$Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$ (z-test, known $\sigma^2$)

$T = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$ (t-test, unknown $\sigma^2$) with $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

$C = \frac{(n-1)S_n^2}{\sigma_0^2}$ ($\chi^2$-test for variance)

| Test type | $H_0$ | $H_1$ | Reject if |
|---|---|---|---|
| One-sided z | $\mu \le \mu_0$ | $\mu > \mu_0$ | $Z > z_\alpha$ |
| Two-sided z | $\mu = \mu_0$ | $\mu \neq \mu_0$ | $\|Z\| > z_{\alpha/2}$ |
| One-sided t | $\mu \le \mu_0$ | $\mu > \mu_0$ | $T > t_{n-1,\alpha}$ |
| Two-sided t | $\mu = \mu_0$ | $\mu \neq \mu_0$ | $\|T\| > t_{n-1,\alpha/2}$ |
| One-sided $\chi^2$ | $\sigma^2 \le \sigma_0^2$ | $\sigma^2 > \sigma_0^2$ | $C > \chi^2_{n-1,\alpha}$ |
| Two-sided $\chi^2$ | $\sigma^2 = \sigma_0^2$ | $\sigma^2 \neq \sigma_0^2$ | $C < \chi^2_{n-1,1-\alpha/2}$ or $C > \chi^2_{n-1,\alpha/2}$ |

## Neyman-Pearson lemma

Let $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$. Then the most powerful test at $\alpha$ rejects $H_0$ if $\frac{L(\theta_1|x)}{L(\theta_0|x)} > k$ for some constant $k$.

1 likelihood ratio $\Lambda\left(\frac{L(\theta_1|x)}{L(\theta_0|x)}\right)$

2 decision rule $\Lambda(x) > k' \Leftrightarrow x < c$, reject if $x < c$

3 type I error probability $P_{\theta_0}(x < c) = \alpha$

4 power (P(correctly rejecting)) Power $= P_{\theta_1}(x < c)$

## confidence intervals

- for mean: known $\sigma^2$: $\bar{X}_n \pm \underbrace{z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{margin of error (ME)}}$ $\qquad n = \left(\frac{z_{\alpha/2}\cdot\sigma}{ME}\right)^2$

  unknown $\sigma^2$: $\bar{X}_n \pm \underbrace{t_{n-1,\alpha/2} \cdot \frac{S_n}{\sqrt{n}}}_{ME}$ $\qquad n = \left(\frac{t_{n-1,\alpha/2}\cdot S_n}{ME}\right)^2$ $z_{\alpha/2}$ also possible

- for proportion: $\hat{p} \pm \underbrace{z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{ME}$ $\qquad$ solving for samplesite $n = \left(\frac{z_{\alpha/2}}{ME}\right)\cdot \hat{p}(1-\hat{p})$

  wanted ME

  use $\hat{p} = 0{,}5$ for worst case