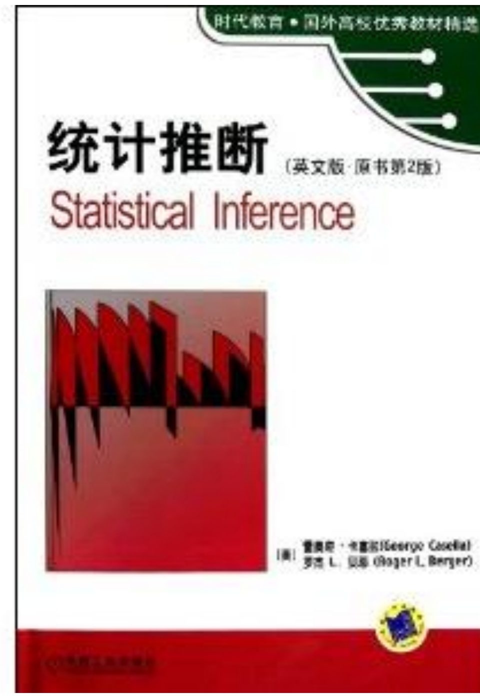
Statistical Inference

Instructor: Li, Caixia

References

- TEXTBOOK: Statistical Inference, 2nd edition; Casella and Berger, 2002.



- Other References:
  - Introduction to Mathematical Statistics, 6th edition; Hogg, McKean and Craig.
  - 高等数理统计(第2版), 茆诗松,王静龙

Chapter 1 ~ 4

- Chapter 1. Probability Theory
- Chapter 2. Transformations and Expectations
- Chapter 3. Common Families of Distributions
- Chapter 4. Multiple Random Vaviables

Chapter 1. Probability Theory

Terminology

- **Sample space**
- **random variable**
- **Distribution functions**
  - **Cumulative distribution function** (cdf)
  - Probability mass function (pmf)
  - Probability density function(pdf)

Chapter 2 Transformations and Expectations

- $Y=g(X)$ and X with cdf $F_X(x)$
- X: discrete random variable, $Y=g(X)$

$$f_Y(y)=P(Y=y)=\sum_{x\in\{x:\, g(x)=y\}} P(X=x) = \sum_{x\in\{x:\, g(x)=y\}} f_X(x)$$

- X: continuous random variable

$$F_Y(y)=P(Y \le y)=P(g(X) \le y)$$
$$=P(\{x:g(x) \le y\})=\int_{\{x:\, g(x)\le y\}} f_X(x)dx$$

- Theorem 2.1.8 $f_x(x)$, $Y=g(X)$ with g is continuous, suppose there exist monotone functions $g_1(x), \ldots, g_k(x)$, defined on $A_1, \ldots, A_k$
- 分段单调

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| \\ 0 \end{cases}$$

Expectation of Random Variable

Let X be a r.v. If X is a discrete or continuous r.v. with pdf $f(x)$ (or pmf p(x) if X is a discrete type r. v.), and

$$\int_{-\infty}^{\infty} |g(x)|\, f(x)dx \; (or \sum_{x\in S} |g(x)|\, p(x))$$

exists. Then the expectation of Y=g(X) is

$$E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx \qquad (or \sum_{x\in S} g(x) p(x))$$

Moments and Moment Generating Functions

- The nth moment, $E(X^n)$. (e.g. $\mu = EX$)
- The nth central moment, $E(X-\mu)^n$,
  - Variance: $\mathrm{Var}(X) = E(X-\mu)^2$
  - Standard deviation: $\sqrt{\mathrm{var}(X)}$
- Moment generating function(mgf),

$$M_X(t) = E\,(e^{tX})$$

$$\frac{d^n}{dt^n} M_X(t)\,|_{t=0} = EX^n \quad \text{e.g.} \quad \frac{d}{dt} M_X(t)\,|_{t=0} = EX$$

Chapter 3 Common Families of Distributions

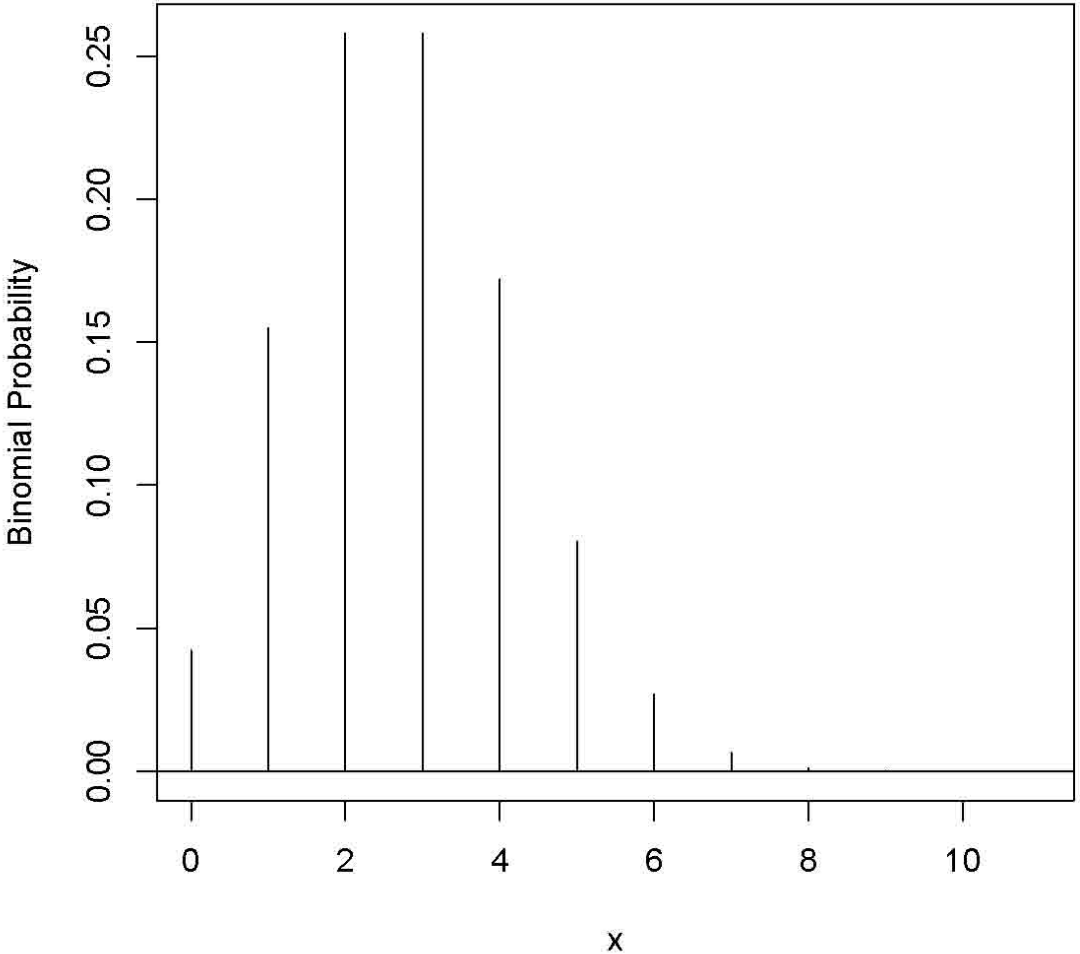
- Discrete Distributions
  - Discrete Uniform
  - Binomial,
  - Poisson,
  - Geometric,
  - Negative binomial (Parskal),
  - Hyper-geometric,
  - ………
- Continuous Distributions
  - Continuous Uniform
  - Exponential
  - Normal
  - Gamma
  - Beta
  - Cauchy
  - Lognormal
  - Double Exponential
  - ……

Discrete Uniform Distribution

$$P(X = x \mid N) = \frac{1}{N}, x = 1, 2, \ldots, N,$$

Binomial Distribution

$$P(X = x \mid n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \cdots, n,$$



the number of observed successes in a sequence of n times Bernoulli trials

Poisson Distribution

$$P(X = x \mid \lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}, x = 0,1,\ldots$$

Negative Binomial Distribution

The number of trials observed until the r-th success occurs in a sequence of Bernoulli trials

$$P(X = x \mid r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, x = r, r+1, \ldots$$

The number of fails observed until the r-th success occurs in a sequence of Bernoulli trials

$$P(X = x \mid r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x = \binom{-r}{x} (-1)^x p^r (1-p)^x, \ x = 0, 1, 2, \cdots$$

Geometric Distribution is a special case of Negative Binomial Distribution with r=1.

$$P(X = x \mid p) = p(1-p)^{x-1}, x = 1, 2, \ldots$$

$$P(X=x|p)=p\,(1-p)^x, \quad x=0, 1, 2, \ldots$$

Hypergeometric Distribution

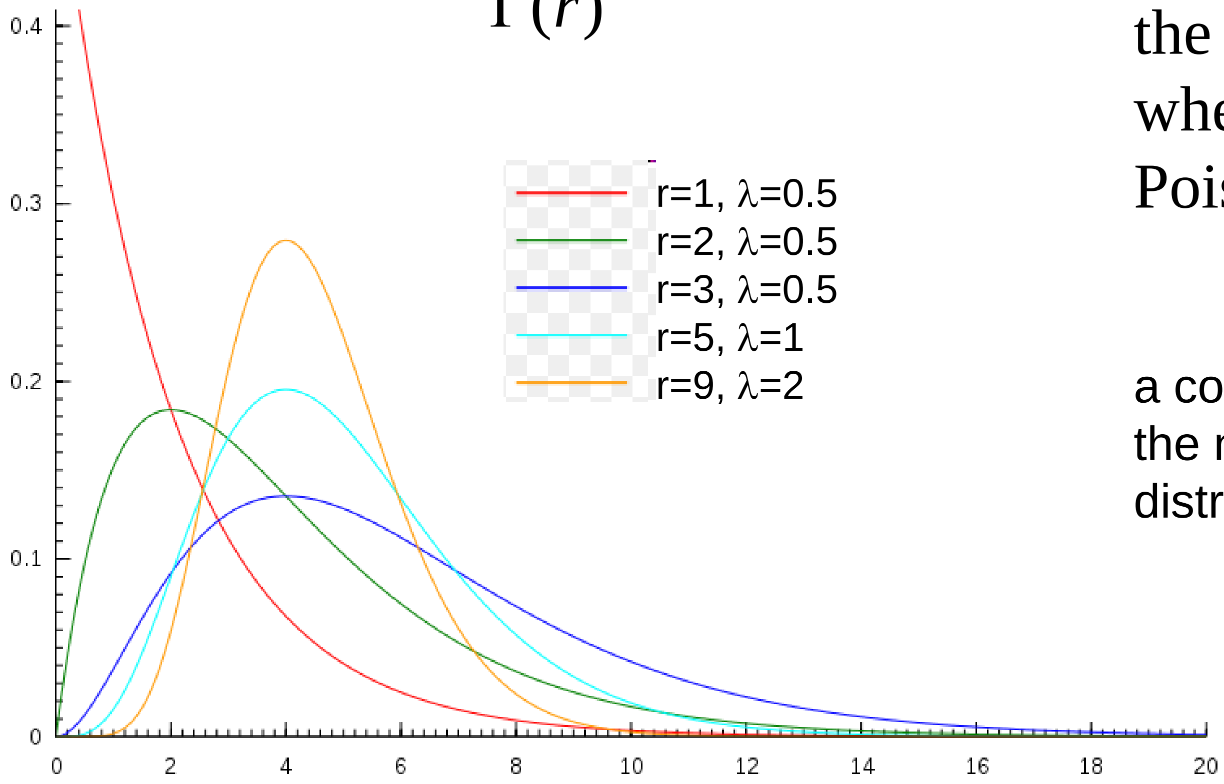
$$P(X = x \mid N, M, K) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{k}}, x = 0, 1, \ldots, K.$$

Continuous Uniform Distribution

$$f(x \mid a,b) = \begin{cases} \dfrac{1}{b-a}, if \;\; x \in (a,b) \\ 0, \qquad otherwise \end{cases}$$

Gamma Distribution

$$f(x \mid r, \lambda) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x \geq 0$$



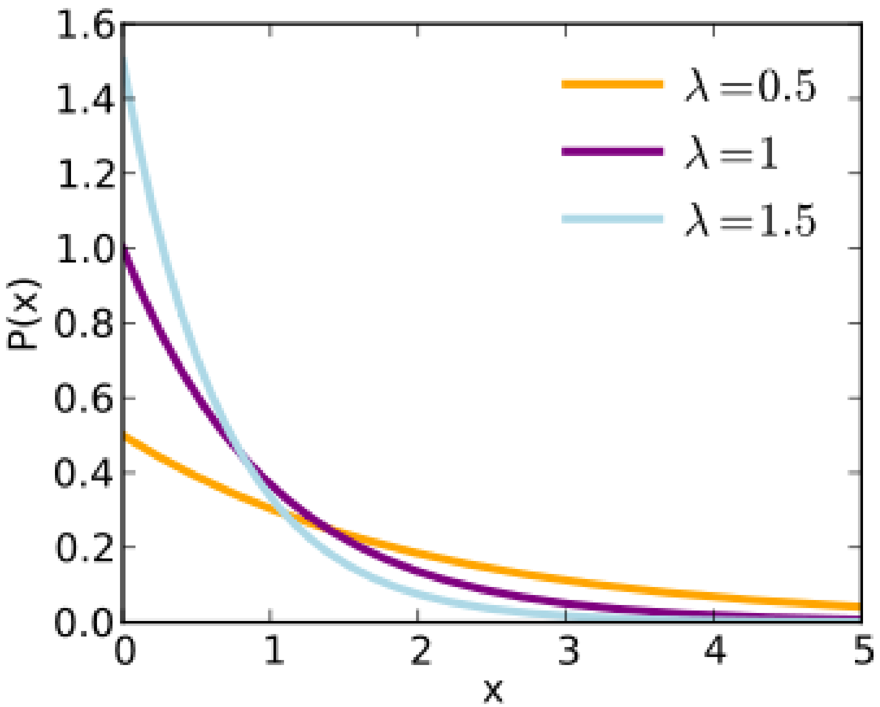
The waiting time until the *r*-th change occurs when observing a Poisson process P(λ).

a continuous counterpart of the negative-binomial distribution

$$\boxed{f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, 0 < x < \infty, \alpha > 0, \beta > 0.}$$

Exponential distribution

$$f(x \mid \lambda) = \lambda e^{-\lambda x}, \quad x \geq 0$$



the waiting time until the first change occurs when observing a Poisson process P(λ).

a continuous counterpart of the geometric distribution

Exponential Distribution is a special case of Gamma Distribution with r=1.

Chi-square Distribution

$$f(x \mid n) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \ \ x \geq 0$$

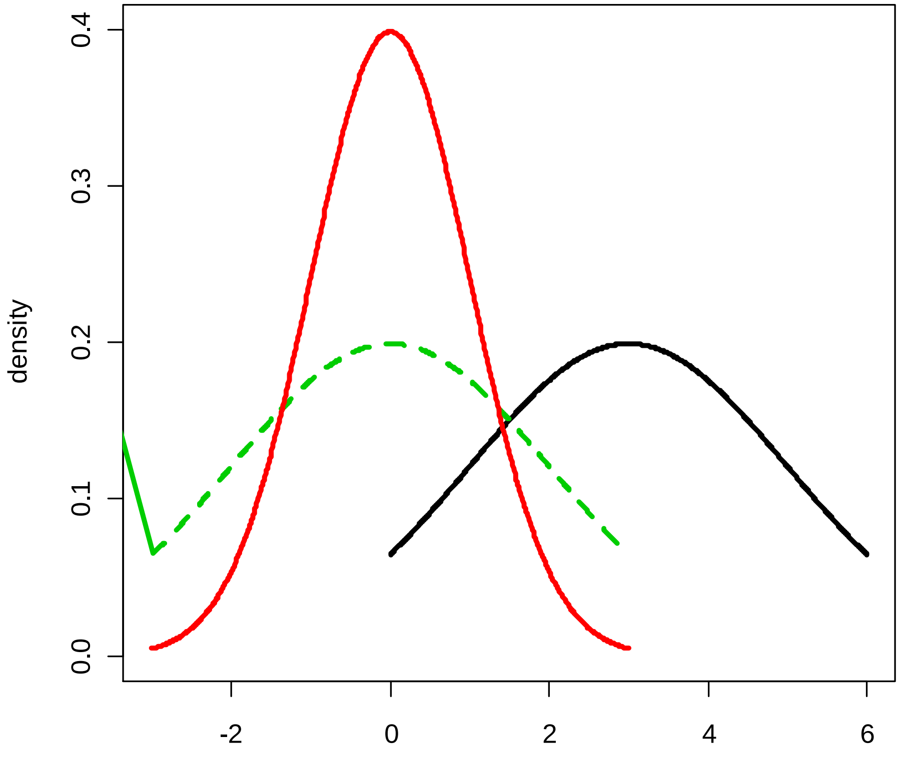
Exponential Distribution is a special case of Gamma Distribution with r=n/2 and λ=1/2 (or α=n/2 and β=2).

Remark: the **sum of the squares** of *k* independent **standard normal** random variables

$$Z_1^2 + Z_2^2 + \cdots + Z_k^2 \sim \chi^2_{(k)}$$

Normal Distribution

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/(2\sigma^2)}, -\infty < x < \infty$$



$$X \sim N\left(\mu, \sigma^2\right)$$

$$\downarrow$$

$$Z = \frac{X - \mu}{\sigma} \sim N\left(0, 1\right)$$

Beta Distribution

$$f(x \mid \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1, \alpha > 0, \beta > 0.$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Remark: Suppose the independent r. vs $X_1 \sim \Gamma(\alpha, \lambda)$, and $X_2 \sim \Gamma(\beta, \lambda)$, then $X_1 + X_2 \sim \Gamma(\alpha+\beta, \lambda)$ is independent $X_1/(X_1 + X_2) \sim \text{Beta}(\alpha, \beta)$.

Cauchy Distribution

$$f(x|\theta)=\frac{1}{\pi}\frac{1}{1+(x-\theta)^2},-\infty<x<\infty,-\infty<\theta<\infty$$

$$E\left|X\right|=\infty$$

Lognormal Distribution

$$f(x\mid\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}e^{-(\log x-\mu)^2/(2\sigma^2)},0<x<\infty,-\infty<\mu<\infty,\sigma>0$$

$$EX=e^{\mu+(\sigma^2/2)}$$

Remark: If X ~ LogN(μ, σ^2), then logX ~ N (μ, σ^2).

Double Exponential Distribution (Laplace Distribution)

$$f(x \mid \mu, \sigma) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0.$$

$$EX = \mu$$

$$VarX = 2\sigma^2$$

Exponential Families

$$f(x|\theta) = h(x)c(\theta)\exp(\sum_{i=1}^{k} w_i(\theta)t_i(x)).$$

- Continuous families
  - Normal, Beta, Gamma, ……
- Discrete families
  - Binomial, Poisson, Negative binomial, ……

Location and Scale Families

$$g(x|\mu) = f\left(x - \mu\right)$$

$$g(x|\sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$$

$$g(x|\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

Chapter 4 Multiple Random Vaviables

- Joint distribution
  - Cdf $F(x,y)=P(X\leq x,Y\leq y)$
  - joint pmf (probability mass function)
  - joint pdf (probability density function)

- Marginal distributions
  - Marginal pmfs of discrete bivariate (X,Y)

  - Marginal pdfs of continuous bivariate

Conditional Distributions

- For discrete bivariate,

  $f(\ y|x\ )=P(\ Y=y \mid X=x\ )=f(x,y)/f_x(x)$

  providing that $f_x(x) > 0$

  $f(\ x|y\ )=P(\ X=x \mid Y=y\ )=f(x,y)/f_Y(y)$

  providing that $f_y(y) > 0$

- For continuous bivariate random vector,

  repalce the pmf function with pdf function

Conditional Expectations

- Expectations

  for discrete bivariate random vector

$$E(g(Y)\,|\,x) = \sum_{y} g(y)\,f(y\,|\,x)$$

  for continuous bivariate random vector

$$E(g(Y)\,|\,x) = \int_{-\infty}^{\infty} g(y)\,f(y\,|\,x)\,dy$$

Conditional Distributions and Independence

- X and Y are called indepent random variables if for every x and y

  $f(x,y)=f_x(x)f_Y(y)$

- if X and Y are indepentdent ,then

  $f(y|x)=f(x,y)/f_x(x)= f_x(x)f_Y(y)/ f_x(x)=f_Y(y)$

- The knowledge X=x does't give us any infomation about Y

Bivariate Transformations

- If (X,Y) is a continuous random vector with joint pdf $f_{X,Y}(x,y)$, and $U=g_1(X,Y)$, $V=g_2(X,Y)$,

Suppose (g_1, g_2): A->B is a one to one transformation

- $x=h_1(u,v)$, $y=h_2(u,v)$, then $f_{U,V}(u,v)=f_{X,Y}(h_1(u,v),h_2(u,v))|J|$

$$|J| = \begin{bmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{bmatrix}$$

- Extending to **not one-to-one transformation**

$$f_{U,V}(u,v) = \sum f_{X,Y}(h_{1i}(u,v),h_{2i}(u,v))|J_i|$$

Covariance and Correlation

- The covariance of X and Y is the number defined by

  $$Cov(X,Y)=E((X-u_X)(Y-u_Y))$$

 The correlation of X and Y is the number defined by

$$\rho_{XY=}Cov(X,Y)/\sigma_X\sigma_Y$$

Covariance and Correlation

- Theorem 4.5.3 For any random variables X and Y, $Cov(X,Y)=EXY-u_X u_Y$

- Theorem 4.5.5 If X and Y are independent random variables , then $Cov(X,Y)=0$ and $\rho_{XY=0}$

- However $Cov(X,Y)=0$ and $\rho_{XY=0}$ does't mean the two are independent(P171)

Covariance and Correlation

- Theorem 4.5.7 For any random variables X and Y,

  a. $-1 \leq \rho_{XY} \leq 1$

  b. $|\rho_{XY}|=1$ if and only if there exist numbers $a \neq 0$ and b such that $P(Y=aX+b)=1$.

  If $\rho_{XY}=1$, then $a>0$, and if $\rho_{XY}=-1$, then $a<0$