Mathematical Statistics and Data Analysis

Lecture 3: Review of Probability - Part II

Lyu Ni

DaSE@ECNU
(lni@dase.ecnu.edu.cn)

September 16, 2019

Outlines

Reading Material

Textbook:

- Rice: Chapter 2;
- Mao: 2.1, 2.4, 2.5;

Random Variable

Definition

A **random variable** (r.v.) X is a function from a sample space Ω into the real numbers, i.e.,

$$X(\omega) = x \in \Re, \forall \omega \in \Omega$$

Remarks

- Use X, Y, Z to represent the random variables and x, y, z to represent the numerical values. E.g. $X = x$ means that random variable X has value x.
- Continuous r.v. or Discrete r.v.: the domain of the function.
  - If the domain is countable, r.v. is a discrete r.v.;
  - If the domain is uncountable, r.v. is a continuous r.v.;

Random Variable

Definition

A **random variable** (r.v.) X is a function from a sample space Ω into the real numbers, i.e.,

$$X(\omega) = x \in \Re, \forall \omega \in \Omega$$

Remarks

- Use X, Y, Z to represent the random variables and x, y, z to represent the numerical values. E.g. $X = x$ means that random variable X has value x.
- Continuous r.v. or Discrete r.v.: the domain of the function.
  - If the domain is countable, r.v. is a discrete r.v.;
  - If the domain is uncountable, r.v. is a continuous r.v.;

Cumulative distribution function

Definition

The **cumulative distribution function** or cdf of a r.v. X is defined to be

$$F(x) = P(X \leq x), -\infty < x < \infty$$

Note that

- X is said to be distributed as $F(x)$, denoted as $X \sim F(x)$.
- Sometimes, $F_X(x)$ is used to be the distribution of X.

Cumulative distribution function

Property

The function $F(x)$ is a cumulative distribution function $\Leftrightarrow$

- **Monotonicity**: $F(x)$ is non-decreasing function of X, i.e. for every $x_1 < x_2$, we have

$$F(x_1) \leq F(x_2).$$

- **Boundedness**: For every x, $0 \leq F(x) \leq 1$ and

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0 \text{ and } F(\infty) = \lim_{x \to \infty} F(x) = 1$$

- **Right-continuousness**: $F(x)$ is a right-continuous, i.e., for every x_0,

$$F_{x \to x_0^+} F(x) = F(x_0) \text{ or } F(x_0 + 0) = F(x_0)$$

Cumulative distribution function

Continuous vs Discrete

- A r.v. X is **continuous** if $F(x)$ is a continuous function;
- A r.v. X is **discrete** if $F(x)$ is a step function.

Identical

Two r.v.s X and Y are **identically distributed** if for every set $A \in \mathcal{F}$,

$$P(X \in A) = P(Y \in A)$$

Theorem

The r.v.s X and Y are identically distributed if and only if $F_X(x) = F_Y(x)$ for every x.

Cumulative distribution function

Example

Toss a fair coin three times. Define r.v.s.

$$X = \# \text{ of heads observed}$$

and

$$Y = \# \text{ of tails observed}$$

We have $P(X = k) = P(Y = k)$ i.e. X and Y are identically distributed. However, we do not have $X(\omega) = Y(\omega)$ for $\omega \in \Omega$.

Note that two r.v.s that are identically distributed are not necessarily equal.

Probability mass function

Definition

For all x, the **probability mass function** (p.m.f.) of a discrete r.v. X on Ω is given by

$$p_i = f(x_i) = P(X = x_i), i = 1, 2, \cdots, n, \cdots$$

For a discrete r.v. X, several presenting method:

- Notations: $X \sim \{p_i\}$;
- Diagram:

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$	$\cdots$

Probability mass function

Property

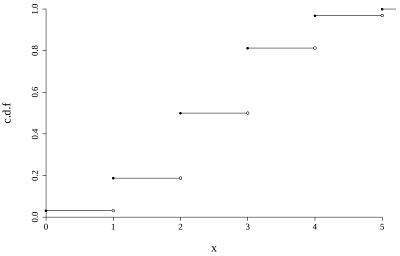
A function $f(x_i)$ is a p.m.f of a discrete r.v. X if and only if

- **Non-negativity**: $f(x_i) \geq 0, i = 1, 2, \cdots;$
- **Normalization**: $\sum_{i=1}^{\infty} f(x_i) = 1$

Formula

The cumulative distribution function (c.d.f.) of the discrete r.v. X is

$$F(x) = \sum_{x_i \leq x} f(x_i)$$



Bernoulli distribution

Definition

A **Bernoulli** random variable takes on only two values: 0 and 1, with probabilities $1-p$ and p, respectively. The p.m.f of X is

$$P(X=x)=\begin{cases} p & x=1 \\ 1-p & x=0 \\ 0 & \text{otherwise} \end{cases}$$

An alternative and useful representation of the p.m.f of X is

$$P(X=x)=\begin{cases} p^x(1-p)^{1-x} & \text{if } x=0 \text{ or } x=1 \\ 0 & \text{otherwise} \end{cases}$$

Bernoulli distribution

If A is an event, then the **indicator random variable**, I_A, takes on the value 1 if A occurs and the value 0 if A does not occur:

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

I_A is a Bernoulli random variable.

Example

- A coin tossing: Head or Tail;
- An exam: Pass or Failure;
- A baby: Male or Female;
- A disease: Cure or Fail;

Binomial distribution

Definition

Suppose that n independent experiments, or trials, are performed, where n is a fixed number, and that each experiment results in a 'success' with probability p and a 'failure' with probability $1-p$. The total number of successes, X, is a **binomial** random variable with parameters n and p.

For the binomial distribution, the p.m.f. of X is

$$P(X=k) = \begin{cases} \binom{n}{k} p^k (1-p)^{1-k}, & k = 0, 1, \cdots, n; \\ 0, & \text{otherwise.} \end{cases}$$

Binomial distribution

Remark

- These probability sum to 1:

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{1-k} = (p + (1-p))^n = 1$$

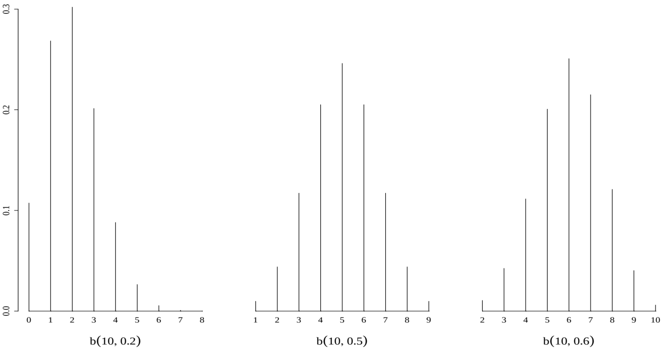
- $X_i \sim B(p)$ and X_is are mutually independent. Then the sum is

$$X = X_1 + X_2 + \cdots + X_n \sim b(n, p)$$

Binomial distribution

Remark (Con'd)

- The shape varies as a function of p.



Poisson distribution

Definition

If a r.v. X is distributed as a **Poisson** distribution with parameter $\lambda(\lambda > 0)$, the p.m.f of X is

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0, 1, 2, \cdots$$

Theorem

A r.v. X is distributed as a binomial distribution with the parameter n and p_n. If $np \to \lambda$ as $n \to \infty$, then the limit distribution of X is Poisson distribution, that is,

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1 - p_n)^{1-k} = \frac{\lambda^n}{k!}e^{-\lambda}$$

Poisson distribution

Definition

If a r.v. X is distributed as a **Poisson** distribution with parameter $\lambda(\lambda > 0)$, the p.m.f of X is

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \cdots$$

Theorem

A r.v. X is distributed as a binomial distribution with the parameter n and p_n. If $np \to \lambda$ as $n \to \infty$, then the limit distribution of X is Poisson distribution, that is,

$$\lim_{n \to \infty} \binom{n}{k} p_n^k (1 - p_n)^{1-k} = \frac{\lambda^n}{k!} e^{-\lambda}$$

Poisson distribution

Solution: Let $\lambda_n = np_n$, that is $p_n = \lambda_n/n$. Then

$$\begin{aligned} p(k) &= \binom{n}{k} p_n^k (1-p_n)^{1-k} \\ &= \frac{n!}{k!(n-k)!}\frac{\lambda_n^k}{n^k}\left(1-\frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{\lambda_n^k}{k!}\frac{n!}{(n-k)!}\frac{1}{n^k}\left(1-\frac{\lambda_n}{n}\right)^{n}\left(1-\frac{\lambda_n}{n}\right)^{-k} \to \frac{\lambda^k}{k!}e^{-\lambda}. \end{aligned}$$

For a fixed k,

$$\lim_{n\to\infty}\lambda_n = \lambda \quad \text{and} \quad \lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k} = e^{-\lambda}$$

$$\lim_{n\to}\frac{n!}{(n-k)!}\frac{1}{n^k} = 1 \quad \text{and} \quad \lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{-k} = 1.$$

Hypergeometric distribution

Definition

Suppose that an urn contains n balls, of which r are black and $n-r$ are white. Let X denote the number of black balls drawn when taking m balls without replacement.

For the **hypergeometric** distribution, the p.m.f. of X

$$P(X=k) = \begin{cases} \dfrac{\dbinom{r}{k}\dbinom{n-r}{m-k}}{\dbinom{n}{m}}, & \text{for } 0 \leq k \leq r \\ 0, & \text{otherwise} \end{cases}$$

Hypergeometric distribution

Remark

- Sample without replacement: hypergeometric distribution;
- Sample with replacement: binomial distribution.
- When $m \ll n$,

$$\frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} \approx \binom{n}{k} p^k (1-p)^{n-k}$$

where $p = r/n$.

Geometric distribution

Definition

The **geometric distribution** is also constructed from independent Bernoulli trials, but from an infinite sequence. On each trial, a success occurs with probability p, and X is the total number of trials up to and including the first success. So that $X = k$, there must be $k-1$ failures followed by a success.
The p.m.f. of X is

$$P(X = k) = \begin{cases} (1-p)^{k-1}p, & \text{for } k = 1, 2, 3, \cdots \\ 0, & \text{otherwise} \end{cases}$$

Property (Memorylessness)

Suppose $X \sim Ge(p)$. For every positive integer m and n, we have

$$P(X > m + n | X > m) = P(X > n)$$

Negative binomial distribution

Definition

The **negative binomial distribution** arises as a generalization of the geometric distribution. Suppose that a sequence of independent trials, each with probability of success p, is performed until there are r successes in all; let X denote the total number of trials.

The p.m.f. of X is

$$P(X=k) = \begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r}, & \text{for } k = 1, 2, 3, \cdots \\ 0, & \text{otherwise} \end{cases}$$

Then, $X \sim Nb(r,p)$.

Negative binomial distribution

Remark

Let

- X_1: the number of trials up to and including the first success;
- X_{k+1}: the number of trails from the kth success up to and including the $(k+1)$th success, $k = 1, 2, \cdots, r-1$.

$$\underbrace{A^c A^c \cdots A^c A}_{X_1} \quad \underbrace{A^c A^c \cdots A^c A}_{X_2} \quad \cdots \quad \underbrace{A^c A^c \cdots A^c A}_{X_r}$$

Namely, if X_i are identically and independently distributed and $X_i \sim Ge(p)$, then

$$X = X_1 + X_2 + \cdots + X_r \sim Nb(r, p)$$

Continuous Random Variable

Definition

The probability density function (p.d.f) $f(x)$ of a continuous r.v. X is the function that satisfies

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t, \text{ for all } x.$$

Property

A function $f(x)$ is a p.d.f of a continuous r.v. X if and only if

- **Non-negativity**: $f(x) \geq 0$;
- **Normalization**: $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$

Notation: 'X has a distribution given by $F(x)$' is abbreviated symbolically by '$X \sim F(x)$' ($X \sim f(x)$).

Continuous Random Variable

Example: Value at Risk (VaR)

Financial firms need to quantify and monitor the risk of their investments. **Value at Risk (VaR)** is a widely used measure of potential losses. It involves two parameters: a time horizon and a level of confidence.

Suppose

- V_0: the current value of the investment;
- V_1: the future value;
- $R = \frac{V_1 - V_0}{V_0}$: the return on the investment;
- $F_R(r)$: the c.d.f. of the return R (a continuous r.v.);

Continuous Random Variable

Example: Value at Risk (VaR) (Con'd)

Let the desired level of confidence be denoted by $1 - \alpha$. We want to find v^*, the VaR. Then

$$\begin{aligned} \alpha &= P(V_0 - V_1 \geq v^*) \\ &= P\left(\frac{V_1 - V_0}{V_0} \leq -\frac{v^*}{V_0}\right) \\ &= F_R\left(-\frac{v^*}{V_0}\right) \end{aligned}$$

Thus, $-\frac{v^*}{V_0}$ is the α quantile, r_α and $v^* = -V_0 r_\alpha$.

The VaR is **minus the current value times the α quantile of the return distribution**.

Normal distribution

- One of the most important distributions in statistics.
- Proposed by Carl Friedrich Gauss.
- A model for measurement errors.

Definition

The density of the normal distribution depends on two parameters μ and σ (where $-\infty < \mu < \infty$, $\sigma > 0$).
For $X \sim N(\mu, \sigma^2)$, the p.d.f. of X is

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

Normal distribution

Remark

- $f(x)$ has a bell-shaped curve with a single peak;
- $f(x)$ is symmetric about μ;
- $f(x)$ has two inflection points at $\mu \pm \sigma$.

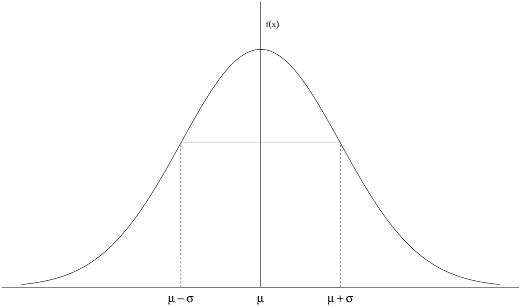
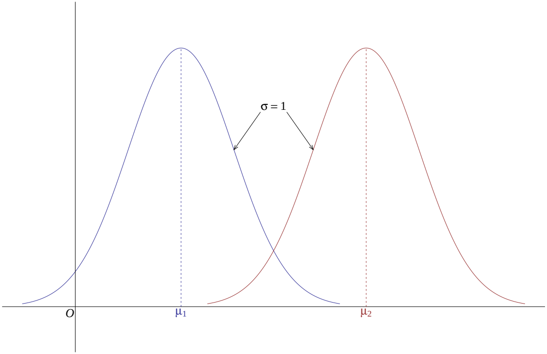


Figure: The p.d.f. of a normal r.v.

Normal distribution

Remark (Con'd)

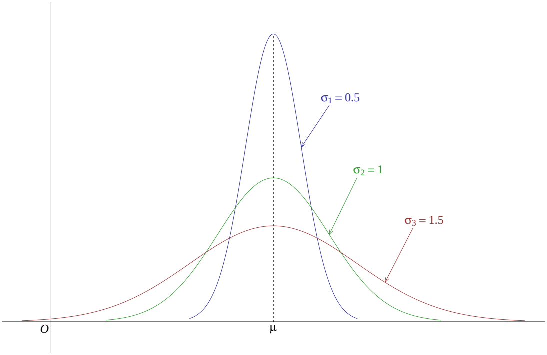
- When σ is fixed, the shape varies as a function of μ.
- μ is a **location** parameter;



Normal distribution

Remark (Con'd)

- When μ is fixed, the shape varies as a function of σ.
- σ is a **scale** parameter;



Normal distribution

Special Case

Suppose a r.v. X is distribution as a normal distribution $N(\mu, \sigma^2)$. If $\mu = 0$ and $\sigma = 1$, then X is said to be a **standard normal** variable.

- U or Z: a standard normal r.v.;
- $\Phi(\cdot)$: the c.d.f. of a standard normal r.v.;
- $\varphi(\cdot)$: the p.d.f. of a standard normal r.v.;

Proposition

If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then $Y \sim N(a\mu + b, a^2\sigma^2)$. Especially, if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Normal distribution

3-σ rule

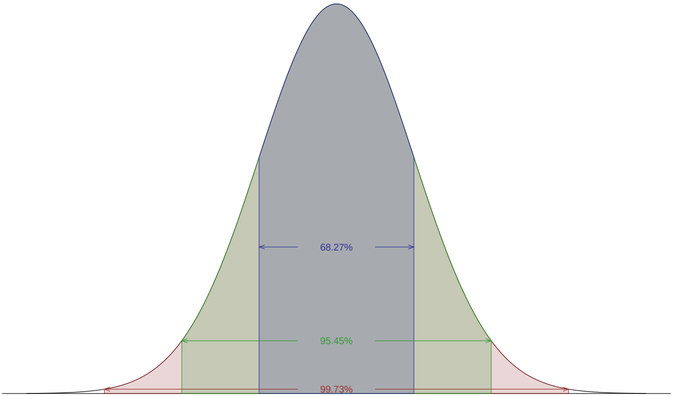
Suppose a r.v. $X \sim N(\mu, \sigma^2)$, then

$$\begin{aligned} P\left(|X-\mu| < k\sigma\right) &= \Phi(k) - \Phi(-k) = 2\Phi(k) - 1 \\ &= \begin{cases} 0.6826, & k = 1; \\ 0.9545, & k = 2, \\ 0.9973, & k = 3; \end{cases} \end{aligned}$$

Applications:

- Control Charts;
- Outlier Detection;

3-σ rule (Con'd)



Uniform distribution

Definition

A uniform r.v. on the interval $[a, b]$ is a model for what we mean when we say 'choose a number at random between a and b.' For X is a **uniform** r.v., the p.d.f. of X is

$$f(x) = \begin{cases} 1/(b-a), & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

From the definition, the c.d.f. of X on $[a, b]$ is

$$F(x) = \begin{cases} 0, & \text{for } x < a \\ x, & \text{for } a \le x < b \\ 1, & \text{for } x \ge b \end{cases}$$

Exponential distribution

Definition

Suppose X is a random variable. X is said to be distributed as an **exponential distribution** if and only if the p.d.f of X is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0; \end{cases}$$

Then, the c.d.f of X is

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0; \\ 0, & \text{otherwise}; \end{cases}$$

- $X \sim Exp(\lambda)$ with a single parameter λ;
- Application: time-to-event;

Exponential distribution

Property (Memorylessness)

Suppose $X \sim Exp(\lambda)$. For every $s > 0$ and $t > 0$, we have

$$P(X > s + t | X > s) = P(X > t)$$

Proof: As we know $P(X > s) = e^{-\lambda s}, s > 0$ since $X \sim Exp(\lambda)$. It is a fact that

$$\{X > s + t\} \subseteq \{X > s\}.$$

Then the conditional probability is

$$P(X > s+t|X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t).$$

Exponential distribution

Example: Exponential vs Poisson

Suppose that $N(t)$ is the number of breakdowns of a machine within a period of time $[0, t]$ and $N(t)$ is distributed as a Poisson distribution with a parameter λt. Let T be the interval time between two successive breakdowns. Then T is an Exponential distribution with the parameter λ.

Solution: Let $N(t) \sim P(\lambda t)$, that is

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k = 0, 1, \cdots$$

Exponential distribution

Example: Exponential vs Poisson (Con'd)

Note that T is a non-negative random variable. The event $\{T \geq t\}$ is equivalent to the event $\{N(t) = 0\}$.

- When $t < 0$, the c.d.f. of T is

$$F_T(t) = P(T \leq t) = 0;$$

- When $t \geq 0$, the c.d.f. of T is

$$\begin{aligned} F_T(t) &= P(T \leq t) = 1 - P(T > t) \\ &= 1 - P(N(t) = 0) = 1 - e^{-\lambda t} \end{aligned}$$

Thus, $T \sim Exp(\lambda)$.

Gamma distribution

Definition

Suppose X is a random variable. X is said to be distributed as an **gamma distribution** if and only if the p.d.f of X is

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0; \end{cases}$$

where the gamma function, $\Gamma(x)$, is defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x, \alpha > 0$$

- $X \sim Ga(\alpha, \lambda)$ with two parameters α and λ;
- α: a **shape** parameter;
- λ: a **scale** parameter;

Gamma distribution

Special Cases

- When $\alpha = 1$, the gamma distribution coincides with the exponential distribution;
- When $\alpha = n/2$ and $\lambda = 1/2$, the gamma distribution coincides with the chi-squared distribution with the degree of freedom n;

Property

- If $X \sim Ga(\alpha, \lambda)$, then $kX \sim Ga(\alpha, \lambda/k)$;
- If $X_1, X_2, \cdots, X_n$ are i.i.d. with a common distribution $Exp(\lambda)$, then $\sum_{i=1}^{n} X_i \sim Ga(n, \lambda)$;

Beta distribution

Definition

Suppose X is a random variable. X is said to be distributed as an **beta distribution** if and only if the p.d.f of X is

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, & 0 \le x \le 1; \\ 0, & \text{otherwise}; \end{cases}$$

- $X \sim Be(a, b)$ with two shape parameters a and b;
- When $a = b = 1$, the beta distribution coincides with the uniform distribution;

Exponential family

Definition

A family of p.d.fs or p.m.fs is said to be an **exponential family** if it has a form as

$$f(x) = h(x) \exp\{\eta(\theta)^\tau T(x) - \zeta(\theta)\}$$

with

- $\theta \in \Theta \subset \Re^d$: a parameter vector;
- $\eta(\theta)$: a function from Θ to $\Re^p$;
- $T(x)$: a random p-vector with a fixed positive integer p;
- $\zeta(\theta)$: is a real-valued function of θ.

Note that it is very helpful to model **heterogeneous** data.

Example: Bernoulli

Suppose X is a random variable. If $X\ B(\theta)$, the p.m.f. of X is

$$\begin{aligned} f(x) &= \theta^x(1-\theta)^{1-x} = \exp\{x\log\theta + (1-x)\log(1-\theta)\} \\ &= \exp\left\{\log\frac{\theta}{1-\theta}\cdot x + \log(1-\theta)\right\} \end{aligned}$$

- $\eta(\theta) = \log\frac{\theta}{1-\theta}$;
- $T(x) = x$;
- $\zeta(\theta) = \log\frac{1}{1-p}$;
- $h(x) = 1$;

Then $f(x) = h(x)\exp\{\eta(\theta)T(x) - \zeta(\theta)\}$. So Bernoulli distribution is one of the exponential families.

Example: Normal

Suppose X is a random variable. If X $N(\mu, \sigma^2)$, the p.d.f. of X is

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \\
&= \exp\left\{-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \log\sqrt{2\pi\sigma^2}\right\}
\end{aligned}
$$

- $\eta(\theta) = \begin{pmatrix} -\frac{1}{2\sigma^2} \\ \frac{\mu}{\sigma^2} \end{pmatrix}$ and $T(x) = \begin{pmatrix} x^2 \\ x \end{pmatrix}$;
- $\zeta(\theta) = \frac{\mu^2}{2\sigma^2} + \log\sqrt{2\pi\sigma^2}$ and $h(x) = 1$;

Then $f(x) = h(x)\exp\{\eta(\theta)T(x) - \zeta(\theta)\}$. So a normal distribution is one of the exponential families.

R code

R code	Description
d + dist.	p.d.f. or p.m.f.
p + dist.	c.d.f.
q + dist.	quantile
r + dist.	random numbers

distribution	dist.	distribution	dist.
Binomial	binom	Normal	norm
Poisson	pois	Exponential	exp
Geometric	geom	Uniform	unif
Negative Binomial	nbinom	Gamma	gamma
Hypergeometric	hyper	Beta	beta