

Lecture Notes 2: Distributions

Professor: Zhihua Zhang

2.1 Distribution Function

The CDF of a discrete random variable X can be expressed as the sum of its probability mass function (pmf) $f_X(x)$ as follows:

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

The CDF of a continuous random variable X can be expressed as the integral of its probability density function (pdf) $f_X(x)$ as follows:

$$F(x) = \int_{-\infty}^x f_X(t) dt$$

and

$$F'(x) = f_X(x)$$

Lemma 2.1 Let F be the CDF for a random variable X , then we have

- (1) $\Pr(X = x) = F(x) - F(x^-)$
- (2) $\Pr(x < X \leq y) = F(y) - F(x)$
- (3) $\Pr(X > x) = 1 - F(x)$
- (4) If X is continuous, then

$$F(b) - F(a) = \Pr(a < X < b) = \Pr(a \leq X < b) = \Pr(a < X \leq b) = \Pr(a \leq X \leq b)$$

Theorem 2.3 Suppose X is a random variable with CDF $F(x)$. The inverse CDF is defined by:

$$F^{-1}(q) = \inf\{x : F(x) > q\}$$

for $q \in [0, 1]$. It's is also called **quantile function**.

Theorem 2.4 The **mode** of a discrete probability distribution is the value at which its pmf takes its maximum value. The mode of a continuous probability distribution is the value x at which its probability density function has its maximum value, so, informally speaking, the mode is at the peak.

Remarks:

- (1) The pmf is always less than or equal to 1, but the pdf can be greater than 1. For example, the uniform distribution on $[0, 1/5]$, the pdf is $f(x) = 5$. The pdf also can be infinite, e.g., $f(x) = \frac{2}{3}x^{-\frac{1}{3}}$.
- (2) $\sum f(x) = 1$ or $\int f(x) = 1$ sometimes is written as $\int dF(x) = 1$ or $\int F(dx) = 1$.
- (3) We call X and Y are equal in distribution iff $F_X(x) = F_Y(x)$ for any x . Notice that it is **not** equal to $X = Y$. For example, $\Pr(X = 1) = \Pr(X = -1) = \frac{1}{2}$. Let $Y = -X$, then X and Y are equal in distribution but $X \neq Y$.

2.2 Discrete Distribution Examples

2.2.1 Uniform Discrete Distribution

Random variable $X \in \{x_1, x_2, \dots, x_n\}$ has a uniform discrete distribution pmf f if

$$f(x) = \begin{cases} \frac{1}{n} & x = x_i, i = 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

2.2.2 Point Mass Distribution

Random variable X has a point mass distribution pmf f if

$$f(x) = \begin{cases} 1 & x = a \\ 0 & \text{otherwise} \end{cases}$$

2.2.3 Bernoulli Distribution

Random variable X has a bernoulli distribution pmf f if

$$f(x) = \begin{cases} p & x = a \\ 1 - p & \text{otherwise} \end{cases}$$

where $p \in [0, 1]$. It can be written as $f(x) = p^x(1 - p)^{1-x}$ also. In binary classification problem, Bernoulli distribution is always used to model the category y .

2.2.4 Poisson Distribution

A discrete random variable X is said to have a Poisson distribution with parameter $\lambda > 0$, if, for $k = 0, 1, 2, \dots$, the probability mass function of X is given by:

$$f(x; \lambda) = \Pr(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \geq 0$$

It is easy to validate $\sum f(x) = 1$ by the Taylor expansion of e^λ .

Remark: If $X_1 \sim \text{Poisson}(\lambda_1)$, $X_2 \sim \text{Poisson}(\lambda_2)$, then $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

2.2.5 Binomial Distribution

A discrete random variable X is said to have a binomial distribution with parameter n and p , we write $X \sim \text{Binomial}(n, p)$. The probability mass function is given by:

$$f(x; n, p) = \Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

for $k = 0, 1, 2, \dots, n$, where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient. It can be interpreted that the probability of exact k successes after n trials.

Remark: If $X_1 \sim \text{Binomial}(n_1, p)$, $X_2 \sim \text{Binomial}(n_2, p)$, then $X_1 + X_2 \sim \text{Binomial}(n_1 + n_2, p)$.

By the way, we introduce something about gamma function and a generalization form of $\binom{n}{k}$.

The gamma function (represented by the capital Greek letter Γ) is an extension of the factorial function, with its argument shifted down by 1, to real and complex numbers. That is, if n is a positive integer:

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

The gamma function is defined for all complex numbers except the negative integers and zero. For complex numbers with a positive real part, it is defined via a convergent improper integral:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

As a generalization of factorial function, $\Gamma(x+1) = x\Gamma(x)$, $\Gamma(1) = 0! = 1$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Also, we can define $\binom{n}{k}$ when n is a real number and k is a integer:

$$\binom{n}{k} = \begin{cases} \frac{n(n-1)\dots(n-k+1)}{k!} & k \geq 0 \\ 0 & k < 0 \end{cases}$$

Then we can get a new binomial theorem: $(1 + z)^n = \sum_k \binom{n}{k} z^k$, $|z| < 1$. It can be proved by Taylor expansion also.

2.2.6 Negative Binomial Distribution

Suppose there is a sequence of independent Bernoulli trials, each trial having two potential outcomes called “success” and “failure”. In each trial the probability of success is p and of failure is $1 - p$. We are observing this sequence until a predefined number r of failures has occurred. Then the random number of successes we have seen, X , will have the negative binomial (or Pascal) distribution:

$$X \sim NB(r, p).$$

The probability mass function of the negative binomial distribution is:

$$f(k; r, p) = \Pr(X = k) = \binom{k+r-1}{k} p^k (1-p)^r$$

for $k = 0, 1, 2, \dots$

Note that

$$\begin{aligned}\binom{r-1}{k} &= \frac{(k+r-1)(k+r-2)\dots r}{k!} \\ &= \frac{(-1)^k(-r)(-r-1)\dots(-r-k+1)}{k!} \\ &= (-1)^k \binom{-r}{k}\end{aligned}$$

Hence,

$$\sum \Pr(X = k) = (1-p)^r \sum (-1)^k \binom{-r}{k} p^k = (1-p)^r (1-p)^{-r} = 1$$

When $r = 1$, the negative binomial distribution is **geometric distribution**: $\Pr(X = k) = (1-p)^{k-1} p$.

Let $p = \frac{\lambda}{\lambda+r}$. If $r \rightarrow \infty$, then $p \rightarrow 0$. We can get Poisson distribution:

$$\begin{aligned}\lim_{r \rightarrow \infty} f(\lambda) &= \lim_{r \rightarrow \infty} \frac{(k+r-1)\dots r}{k!} \left(\frac{\lambda}{r+\lambda}\right)^k \left(\frac{r}{r+\lambda}\right)^r \\ &= \lim_{r \rightarrow \infty} \lambda^k \frac{(k+r-1)\dots r}{k!} \left(\frac{1}{r+\lambda}\right)^k \left(\frac{1}{\frac{\lambda}{r}+1}\right)^r \\ &= \lim_{r \rightarrow \infty} \frac{\lambda^k}{k!} \frac{(k+r-1)\dots r}{(\lambda+r)^k} \frac{1}{\left(1+\frac{\lambda}{r}\right)^r} \\ &= \frac{\lambda^k}{k!} e^{-\lambda}\end{aligned}$$

Bernoulli Distribution and Measure Let $\Omega = [0, 1]$, $P([a, b]) = b - a$, $0 \leq a \leq b \leq 1$ (Lebesgue measure). Fix $P \in (0, 1)$ and let

$$X(\omega) = \begin{cases} 1 & \omega \leq p \\ 0 & \omega > p \end{cases}$$

Hence, $\Pr(X = 1) = \Pr(\omega \leq p) = \Pr([0, p]) = p$, $\Pr(X = 0) = \Pr(\omega > p) = \Pr((p, 1]) = 1 - p$.

HomeWork:

(1) If $\lim_{n \rightarrow \infty} a_n = a$, show that $\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$.

(2) Prove the Stirling Formula.

$$\begin{aligned}\lim_{p \rightarrow \infty} \frac{\ln \Gamma(p)}{\frac{1}{2} \ln(2\pi) + (p - \frac{1}{2}) \ln p - p} &= 1 \\ \lim_{p \rightarrow \infty} \frac{\Gamma(p)}{(2\pi)^{\frac{1}{2}} p^{p-\frac{1}{2}} e^{-p}} &= 1\end{aligned}$$

2.3 Continuous Distribution Examples

2.3.1 Continuous Uniform Distribution

A continuous random variable X is said to have a uniform distribution in $[a, b]$, if the probability density function is given by:

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

2.3.2 Normal(Gaussian) Distribution

A continuous random variable X is said to have a Gaussian distribution with parameter μ and σ , if the probability density function of X is given by:

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

denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$. The cumulative distribution function of Gaussian random variable X with parameter $\mu = 0$ and $\sigma = 1$ ($X \sim \mathcal{N}(0, 1)$) is:

$$\Phi(z) = \Pr(X < z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

2.3.3 Dirac Distribution

The Dirac function, or δ function can be loosely thought of as a function on the real line which is zero everywhere except at the origin, where it is infinite,

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

and which is also constrained to satisfy the identity

$$\int_{-\infty}^{\infty} \delta(x) dx = 1$$

2.3.4 Exponential Power Distribution

$$f(x) = \frac{1}{2^{\frac{q+1}{q}} \Gamma(\frac{q+1}{q}) \sigma} e^{(-\frac{1}{2})^{\frac{q+1}{q}} \left(\frac{x-\mu}{\sigma}\right)^q}$$

This family includes the normal distribution when $q = 2$ and it includes the Laplace distribution when $q = 1$: $f(x) = \frac{1}{4\sigma} e^{-\frac{|x-\mu|}{2\sigma}}$

To validate $\int f(x) = 1$, the following formulas may help. For $a > 0, p > 0$,

$$\begin{aligned}\int_0^\infty x^{p-1} e^{-ax} dx &= a^{-p} \Gamma(p) \\ \int_0^\infty x^{-(p+1)} e^{-ax^{-1}} dx &= a^{-p} \Gamma(p) \\ \int_0^\infty x^{p-1} e^{-ax^2} dx &= \frac{1}{2} a^{-\frac{p}{2}} \Gamma\left(\frac{p}{2}\right) \\ \int_0^\infty x^{-(p+1)} e^{-ax^{-2}} dx &= \frac{1}{2} a^{-\frac{p}{2}} \Gamma\left(\frac{p}{2}\right)\end{aligned}$$

More generally, for $a > 0, p > 0$,

$$\begin{aligned}\int_0^\infty x^{p-1} e^{-ax^q} dx &= \frac{1}{q} a^{-\frac{p}{q}} \Gamma\left(\frac{p}{q}\right) \\ \int_0^\infty x^{-(p+1)} e^{-ax^{-q}} dx &= \frac{1}{q} a^{-\frac{p}{q}} \Gamma\left(\frac{p}{q}\right)\end{aligned}$$

2.3.5 Generalized Inverse Gaussian Distribution

A continuous random variable X is said to have generalized inverse Gaussian distribution (GIG) with parameters α, β, r , if the probability density function of X is given by:

$$f(x) = \frac{(\alpha/\beta)^{r/2}}{2K_r(\sqrt{\alpha\beta})} x^{r-1} e^{-(\alpha x + \beta/x)/2}, x > 0$$

where K_r is a modified Bessel function of third kind with index $r, \alpha > 0, \beta > 0$.

2.3.6 Chi-Squared Distribution

A continuous random variable X is said to have chi-squared distribution, if the probability density function of X is given by:

$$f(x) = \frac{1}{\Gamma(\frac{p}{2}) 2^{\frac{p}{2}}} x^{\frac{p}{2}-1} e^{-\frac{x}{2}}$$

2.3.7 Beta Distribution

A continuous random variable X is said to have beta distribution, if the probability density function of X is given by:

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$