Lecture 29: Nov 3

Last time

- Common Continuous Distribution

Today

- Common Continuous Distribution

Common continuous distributions

Normal Distribution Introduced by De Moivre (1667 - 1754) in 1733 as an approximation to the binomial. Later studied by Laplace and others as part of the Central Limit Theorem. Gauss derived the normal as a suitable distribution for outcomes that could be thought of as sums of many small deviations.

- Sample space: $\mathbb{R} = (-\infty, \infty)$
- pdf: For $Y \sim N(\mu, \sigma^2)$,

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

- cdf: There is no closed form.
- When $\mu = 0$ and $\sigma = 1$, the distribution is called *standard normal*:

$$\Phi(y) = \Pr(Y \leqslant y), \quad \Phi(-y) = 1 - \Phi(y)$$

- Mean:

$$EY = \mu$$

- Variance:

$$Var(Y) = E(Y-\mu)^2 = \sigma^2$$

- Higher central moments:

$$E(Y-\mu)^m = \begin{cases} \frac{m!}{2^{m/2}(m/2)!}\sigma^m & m \text{ is even} \\ 0 & m \text{ is odd} \end{cases}$$

- In particular:

$$\mu_3 = E(Y-\mu)^3 = 0 (\text{Skewness})$$
$$\mu_4 = E(Y-\mu)^4 = 3\sigma^4$$

- Moment generating function:

$$M_Y(t) = \exp(\mu t + \sigma^2 t^2/2)$$

Standardization

$$Y \sim N(\mu, \sigma^2) \iff Z = \frac{Y - \mu}{\sigma} \sim N(0,1)$$

Shifting and scaling:

$$Z \sim N(0,1) \iff Y = \sigma Z + \mu \sim N(\mu, \sigma^2)$$

Notes

- Normal distribution is useful in many practical settings. E.g. measurement error.
- Plays an important role in *sampling distributions* in *large samples*, since the Central Limit Theorem syas that the sums of independent identically distributed random variables are approximately normal
- There are many important distributions that can be derived from functions of normal random variables (e.g. χ^2, t, F). We will briefly present the pdf's and sample spaces of these distributions.

χ^2 distribution If $Z \sim N(0,1)$, then $X = Z^2$ has the χ^2 distribution with 1 degree of freedom. More generally, we have the χ^2 distribution with v degrees of freedom with pdf:

$$f(x) = \frac{(x/2)^{\frac{v}{2}-1} e^{-x/2}}{2\Gamma(v/2)}, \quad x > 0$$

where $\Gamma(a)$ is the complete gamma function,

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

The $\chi^2(v)$ distribution is a special case of the gamma distribution, so it is easier to derive its properties from the gamma.

Facts about the Gamma function

- $\Gamma(a+1) = a\Gamma(a), a > 0$
- $\Gamma(1) = 1$
- $\Gamma(n) = (n-1)!$
- $\Gamma(1/2) = \sqrt{\pi}$

Student's t and F distributions Y has a t_k distribution (t with k degrees of freedom) if its pdf can be written as:

$$f(y) = \frac{\Gamma\left[(v+1)/2\right]}{\sqrt{v\pi}\,\Gamma(v/2)} \frac{1}{(1 + y^2/v)^{(v+1)/2}}, \quad -\infty < y < \infty$$

Y has an $F(v_1, v_2)$ distribution if its pdf can be written as:

$$f(y) = \frac{(v_1/v_2)\Gamma\left[(v_1+v_2)/2\right](v_1 y/v_2)^{v_1/2-1}}{\Gamma(v_1/2)\Gamma(v_2/2)(1+v_1 y/v_2)^{(v_1+v_2)/2}}, \quad 0 \leqslant y < \infty$$

There are many important properties and relationships between these three distributions (e.g., χ^2_k is the distribution of the sum of the squares of k independent standard normals).

Gamma distribution Notation: $Y \sim Gamma(a, \lambda)$.

- pdf:

$$f(y) = \frac{\lambda e^{-\lambda y}(\lambda y)^{a-1}}{\Gamma(a)}, \quad y \geqslant 0$$

  where $\Gamma(a)$ is the gamma function,

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

- cdf: In general, there is no closed form, unless a is an integer.
- moments:

$$E(Y) = a/\lambda$$
$$Var(Y) = a/\lambda^2$$

- MGF:

$$M_Y(t) = \left(\frac{1}{1-t/\lambda}\right)^a, \quad t < \theta$$

Another parameterization Same as the exponential distribution, we can let $\beta = \frac{1}{\lambda}$, then we have

- pdf:

$$f(y) = \frac{y^{a-1} e^{-y/\beta}}{\Gamma(a)\beta^a}, \quad y \geqslant 0$$

- moments:

$$EX = \alpha\beta$$
$$Var(X) = \alpha\beta^2$$

- MGF:

$$M_Y(t) = \left(\frac{1}{1-t\beta}\right)^a, \quad t < \frac{1}{\beta}$$

Notes:

- The special case $a = 1$ corresponds to an $exponential(\lambda)$

- The parameter a is known as the *shape parameter*, since it most influences the peakedness of the distribution.
- The parameter β is called the *scale parameter* since most of its influence is on the spread of the distribution.
- The special case $Gamma(a = n/2, \lambda = 1/2)$, for integer n, corresponds to the χ^2_n distribution with n degrees of freedom.
- The gamma distribution can be derived as the sum of a independent $exponential(\lambda)$ distributions.

Beta distribution Notation: $Y \sim Beta(a, b)$.

- Sample space: $[0, 1]$
- pdf:

$$f(y) = \frac{y^{a-1}(1-y)^{b-1}}{B(a,b)}, \quad 0 \leqslant y \leqslant 1$$

  where $B(a, b)$ is the Beta function,

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

  and $\Gamma(a)$ is the gamma function. Note that if a and b are integers, then $B(a, b)$ can be calculated in closed form.
- cdf: In general, there is no closed form, except if a and b are integers.
- moments:

$$EY = \frac{a}{a+b}$$
$$Var(Y) = \frac{ab}{(a+b)^2(a+b+1)}$$

  The beta distribution is very flexible, and can take a wide variety of shapes by varying its parameters.
- Special case: $Beta(1, 1) = U(0, 1)$.

Omitted distributions: Weibull distribution, and Cauchy distribution.

Exponential Families A family of pdfs or pmfs with vector parameter $\boldsymbol{\theta}$ is called an *exponential family* if it can be expressed as

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta})t_j(x)\right), \quad x \in S \subset \mathbb{R} \tag{1}$$

where S is not defined in terms of $\boldsymbol{\theta}$, $h(x)$, $c(\boldsymbol{\theta}) \geqslant 0$ and the functions are just functions of the parameters specified; i.e. h is free of $\boldsymbol{\theta}$, $c(\boldsymbol{\theta})$ is free of x, etc...

Examples:

- One-dimensional: Exponential, Poisson
- Two-dimensional: Gaussian

Exponential family parameterizations are unique except for multiplying constant factors.

Example: Gaussian Let $f(x|\mu, \sigma^2)$ be the $n(\mu, \sigma^2)$ family of pdfs, where $\boldsymbol{\theta} = (\mu,$ $sigma)$. Then

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

Thus

$$\begin{aligned} h(x) &= \tfrac{1}{\sqrt{2\pi}} & c(\mu, \sigma) &= \tfrac{1}{\sigma} \exp\left(-\tfrac{\mu^2}{2\sigma^2}\right) \\ w_1(\mu, \sigma) &= -\tfrac{1}{2\sigma^2} & w_2(\mu, \sigma) &= \tfrac{\mu}{\sigma^2} \\ t_1(x) &= x^2 & t_2(x) &= x \end{aligned}$$

The parameter space is $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$.

Example: Binomial Let $f(x|p)$ be the $binomial(n, p)$, $0 < p < 1$ family of pmfs.

$$\begin{aligned} f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} (1-p)^n \left[\frac{p}{1-p}\right]^x \\ &= \binom{n}{x} (1-p)^n \exp\left[\log\left(\frac{p}{1-p}\right) x\right] \end{aligned}$$

Thus,

$$\begin{aligned} h(x) &= \tbinom{n}{x}, \quad x = 0, \ldots, n & w_1(p) &= \log\left(\tfrac{p}{1-p}\right) \\ c(p) &= (1-p)^n, 0 < p < 1 & t_1(x) &= x \end{aligned}$$

Note that this works when p is considered the parameter, while n is fixed. Also, p cannot be 0 or 1. Otherwise, the range changes.

More examples The following distributions belong to Exponential families:

- Continuous: exponential, Gaussian, gamma, beta, χ^2
- Discrete: Poisson, geometric, binomial (fixed # trials), negative binomial (fixed # successes)

The following distributions not exponential families:

- Continuous: t, F, unifrom E.g.:$X \sim U(0, \theta)$

$$f_X(x) = \theta^{-1} 1(0 < x < \theta)$$

- Discrete: uniform, hypergeometric

Theorem If X is a random variable with pdf or pmf of the form 1, then

$$E\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) = -\frac{\partial}{\partial \theta_j} \log c(\boldsymbol{\theta})$$
$$Var\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) = -\frac{\partial^2}{\partial \theta_j^2} \log c(\boldsymbol{\theta}) - E\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2} t_i(X)\right).$$

Although these equations may look formidable, when applied to specific cases they can work out quite nicely. Their advantage is that we can replace integration or summation by differentiation, which is often more straightforward.

Example (Normal exponential family) Let $f(x|\mu, \sigma^2)$ be the $N(\mu, \sigma^2)$ family of pdfs, where $\boldsymbol{\theta} = (\mu, \sigma), -\infty < \mu < \infty, \sigma > 0$. Then

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

Define

$$\theta_1 = \tfrac{1}{\sigma^2} > 0, \quad \theta_2 = \tfrac{\mu}{\sigma^2} \in \mathbb{R}$$

Then

$$f_X(x) = \frac{\sqrt{\theta_1}}{\sqrt{2\pi}} \exp\left(-\frac{\theta_2^2}{2\theta_1}\right) \exp\left(-\theta_1 \frac{x^2}{2} + \theta_2 x\right)$$

and

$$\begin{array}{ll} h(x) = 1 \text{ for all } x; & \\ c(\boldsymbol{\theta}) = c(\theta_1, \theta_2) = \exp\left(-\frac{\theta_2^2}{2\theta_1}\right), & (\theta_1, \theta_2) \in (0, \infty) \times \mathbb{R} \\ w_1(\boldsymbol{\theta}) = \theta_1 & t_1(x) = -x^2/2 \\ w_2(\boldsymbol{\theta}) = \theta_2 & t_2(x) = x \end{array}$$

Therefore, by the above theorem

$$\begin{aligned} E(X) &= -\frac{\partial}{\partial \theta_2} \log c(\boldsymbol{\theta}) = \frac{\theta_2}{\theta_1} = \mu \\ Var(X) &= -\frac{\partial^2}{\partial \theta_2^2} \log c(\boldsymbol{\theta}) = -\frac{1}{\theta_1} = \sigma^2 \end{aligned} \tag{2}$$