BIOS 660/BIOS 672 (3 Credits):
Probability and Statistical Inference I

Jianwen Cai

`https://sakai.unc.edu/portal/site/bios660-bios672-3-credits`

Notes 12

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: $R = (-\infty, \infty)$

pdf: For $Y \sim N(\mu, \sigma^2)$,

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

If $\mu = 0$ and $\sigma^2 = 1$ is referred to as the *standard normal*.

cdf: There is no closed form. The notation $\Phi(x)$ is often used for $F(x) = P(Y \leq x)$ for the standard normal case. Many books have tables of its values for $x > 0$. Values for $x < 0$ can be obtained by the formula $\Phi(-x) = 1 - \Phi(x)$.

Normal Moments

Mean:

$$\mathsf{E}Y = \mu$$

Variance:

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

Higher central moments:

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

In particular:

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

Standartization

Standartization:

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

Shifting and scaling:

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

Easy to prove using the mgf.

Density integrates to 1

Theorem:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy = 1$$

Proof:
This is not as easy as one might think. Call the integral I. Then,

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(x^2+y^2)}{2}} dx dy \end{aligned}$$

Now make a change of variables to polar coordinates, i.e. put

$$y = r \sin\theta \quad x = r\cos\theta, \quad 0 < \theta \le 2\pi, \quad 0 < r < \infty$$

cont.

Now, $dxdy \rightarrow rdrd\theta$, because

$$\text{Jacobian} = J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r$$

$$\begin{aligned} I^2 &= \int_{r=0}^{\infty}\int_{\theta=0}^{2\pi} \frac{e^{-r^2/2}}{2\pi} rdrd\theta \\ &= \int_0^{\infty} e^{-r^2/2} rdr = -e^{-r^2/2}\Big|_0^{\infty} = 0 - (-1) = 1 \end{aligned}$$

Notes

- Normal distribution useful in many practical settings
- Plays an important role in *sampling distributions in large samples*, since the Central Limit theorem says that 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 see much more on this later, for now, 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 ν degrees of freedom with pdf:

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

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

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

Note that if a is an integer, $\Gamma(a) = (a-1)!$.

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

Student's t and F distributions

Y has a t_k distribution (t with ν degrees of freedom) if its *pdf* can be written as:

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

Y has an $F(\nu_1, \nu_2)$ distribution if its *pdf* can be written as:

$$f(y) = \frac{(\nu_1/\nu_2)\Gamma[(\nu_1+\nu_2)/2)](\nu_1 y/\nu_2)^{\nu_1/2-1}}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)(1+\nu_1 y/\nu_2)^{(\nu_1+\nu_2)/2}}, \qquad 0 \leq y < \infty$$

There are many important properties and relationships between these three distributions (e.g. χ_k^2 is the distribution of the sum of the squares of k independent standard normals). We'll come back to these in a few weeks when we do *sampling distributions and transformations of the normal distribution*.

Gamma distribution

Notation: $Y \sim gamma(a, \lambda)$.
pdf:

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

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

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

Note that if a is an integer, $\Gamma(a) = (a-1)!$.

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

moments

$$\begin{aligned} \mathrm{E}(Y) &= a/\lambda \\ \mathrm{Var}(Y) &= a/\lambda^2 \end{aligned}$$

Notes

- The special case $a = 1$ corresponds to an exponential(λ)
- Can be thought of as a flexible generalization of the exponential (a can be interpreted as a *shape parameter*)
- The special case $gamma(n/2, 1/2)$, for integer n, corresponds to the χ^2 distribution with n degrees of freedom.
- We will see later in the class that the gamma distribution can be derived as the sum of a independent exponential(λ) distributions
- When a is an integer, the *gamma*(a, λ) distribution can be derived as the distribution of time until the occurence of the a^{th} event in a Poisson process.

Weibull distribution

This is another useful generalization of the exponential. It is useful to begin with the *cdf* instead of the *pdf*:
sample space: $[v, \infty]$
cdf:

$$F(y) = 1 - \exp\left[-\left(\frac{y-v}{\alpha}\right)^{\beta}\right], \qquad y \geq v$$

It follows that the *pdf* is:

$$f(y) = \frac{\beta}{\alpha}\left(\frac{y-v}{\alpha}\right)^{(\beta-1)} \exp\left[-\left(\frac{y-v}{\alpha}\right)^{\beta}\right], \qquad y \geq v$$

The usual case is $v = 0$.
If $\beta = 1$ we get an exponential with parameter $\lambda = 1/\alpha$.

Cauchy distribution

This is a famous distribution to mathematical statisticians, since it often serves as a useful counterexample.
pdf

$$f(y) = \frac{1}{\pi}\frac{1}{[1+(y-\mu)^2/\sigma^2]} \quad \text{for } -\infty < y < \infty$$

The Cauchy with $\mu = 0$, $\sigma = 1$, corresponds to the t-distribution with 1 degree of freedom.

While the moments of the Cauchy are not defined, its quantiles are (HW).

The Cauchy is not just a pathological case. We'll see later that the ratio of two standard normals is Cauchy. So ratios of observations can be problematic (e.g. BMI).

Beta distribution

Notation: $Y \sim beta(a,b)$.
sample space: [0,1]
pdf:

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

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

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

where $\Gamma(a)$ is the complete gamma function. The normalizing constant is required so that $\int_0^1 f(x)dx = 1$.
Note that if a and b are integers, then $B(a,b)$ can be calculated in closed form.

cont.

cdf: In general, there is no closed form, except if a and b are integers.

moments

$$\begin{aligned} \mathrm{E}(Y) &= \frac{a}{a+b} \\ \mathrm{Var}(Y) &= \frac{ab}{(a+b)^2(a+b+1)} \end{aligned}$$

The beta distribution is very flexible, and can take a wide variety of shapes by varying its parameters.

Special case: $beta(1,1) = U(0,1)$.

* Read C-B Section 3.3 (normal, beta, Cauchy, lognormal and double exponential)

Location and Scale families

Let $f(x)$ be any pdf. Then the family of pdfs

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

for $\mu \in \mathbb{R}$, $\sigma > 0$, is called a location-scale family.
If $\mu = 0$ we get a scale family; if $\sigma = 1$ we get a location family.

Examples: Normal, Laplace, Cauchy, exponential.

Properties: Let $Z \sim f(z)$ and $X = \sigma Z + \mu$. Then

1. X has pdf $f_{\mu,\sigma}(x)$.
2. $$\mathsf{E}(X) = \sigma \mathsf{E}(Z) + \mu, \qquad \mathsf{Var}(X) = \sigma^2 \mathsf{Var}(Z)$$

Group families

Let $\mathcal{G}$ be a class of 1-to-1 functions $g : \mathbb{R} \to \mathbb{R}$. The class of transformations $\mathcal{G}$ is called a *transformation group* if

1. $\mathcal{G}$ is closed under composition: $g_1, g_2 \in \mathcal{G}$ implies $g_2 \cdot g_1 \in \mathcal{G}$.
2. $\mathcal{G}$ is closed under inversion: $g \in \mathcal{G}$ implies $g^{-1} \in \mathcal{G}$.

Given a rv Z with cdf $F(z)$, the class

$$\{X = g(Z), g \in \mathcal{G}\}$$

is a group family.

Group families: Examples

- *Parametric*: Location-scale families, $g(z) = \sigma z + \mu$, $\sigma > 0$, $\mu \in \mathbb{R}$.
- *Non-parametric*: Let $\mathcal{G}$ is the class of all continuous strictly increasing functions $g(z)$ such that

$$\lim_{z \to -\infty} g(z) = -\infty, \qquad \lim_{z \to \infty} g(z) = \infty$$

  Let Z be a rv supported on $(-\infty, \infty)$. Then the class $\{X = g(Z), g \in \mathcal{G}\}$ is the class of all rvs suported on $(-\infty, \infty)$ whose cdfs are continuous and strictly increasing.
- *Non-parametric*: Same as before with the additional restriction that Z has a symmetric distribution about 0 and g is odd: $g(-z) = -g(z)$. The generated rvs are now the class of all rvs with symmetric distributions about 0.