1 Common Families of Distributions

Lots of this chapter is standard definitions of distributions, so is omitted

1.1 Some Distributions

Definition 1 (Normal Distribution)**.** If $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then $\boldsymbol{X}$ has pdf

$$f_{\boldsymbol{X}}(x_1, \dots, x_k | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right)$$

.

1.1.1 Chi-Squared Distribution

Definition 2. The *chi-squared distribution with* p *degrees of freedom* has pdf

$$\chi_p^2 \sim \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad 0 < x < \infty.$$

Theorem 1 (Some facts)**.**

a. If $Z \sim \mathcal{N}(0, 1)$ then $Z^2 \sim \chi_1^2$

b. If $X_1, \dots, X_n$ are independent $X_i \sim \chi_{p_i}^2$ then $X_1 + \dots + X_n \sim \chi_{p_1 + \dots + p_n}^2$.

1.1.2 Student's t-Distribution

Definition 3 (Student's t-distribution)**.** $T \sim t_p$, a *t-distribution with p degrees of freedom* if it has pdf

$$f_T(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\sqrt{p\pi}} \frac{1}{(1 + t^2/p)^{(p+1)/2}}, \quad t \in \mathbb{R}$$

If $p = 1$ then this is the Cauchy distribution.

Remark 1. If $X_1, \dots, X_n$ are a random sample from $\mathcal{N}(\mu, \sigma^2)$ then

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

This is often taken as the definition. Note that the denominator is independent of the numerator.

Lemma 1 (Moments and mgf of t-distribution)**.** Student's t has no mgf because it does not have moments of all orders: t_p has only $p - 1$ moments. If $T_p \sim t_p$ then

$$\mathbb{E}[T_p] = 0 \quad p > 1$$
$$\operatorname{Var}[T_p] = \frac{p}{p-2} \quad p > 2$$

1.1.3 Snedcor's F-Distribution

Definition 4 (Snedcor's F-distribution)**.** A random variable $X \sim F_{p,q}$ has *F-distribution with p and q degrees of freedom* if its pdf is

$$f_X(x) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} (p/q)^{p/2} \frac{x^{(p/2)-1}}{(1 + px/q)^{(p+q)/2}}, \quad 0 < x < \infty.$$

Remark 2. If $X_1, \dots, X_n$ is a random sample from $\mathcal{N}(\mu_X, \sigma_X^2)$ and $Y_1, \dots, Y_m$ is an independent random sample from $\mathcal{N}(\mu_Y, \sigma_Y^2)$, then

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1,m-1}.$$

This is often taken as the definition.

Theorem 2 (Some facts)**.**

a. $X \sim F_{p,q} \implies 1/X \sim F_{q,p}$

b. $X \sim t_q \implies X^2 \sim F_{1,q}$

c. $X \sim F_{p,q} \implies \frac{(p/q)X}{1+(p/q)X} \sim \text{beta}(p/2, q/2)$

1.1.4 Multinomial Distribution

Definition 5 (Multinomial Distribution)**.** Let m and n be positive integers and let $p_1, \dots, p_n \in [0,1]$ satisfy $\sum_{i=1}^n p_i = 1$. Then the random vector $(X_1, \dots, X_n)$ has *multinomial distribution with m trials and cell probabilities* $p_1, \dots, p_n$ if the joint pmf of $(X_1, \dots, X_n)$ is

$$f(x_1, \dots, x_n) = \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n} = m! \prod_{i=1}^n \frac{p_i^{x_i}}{x_i!}$$

on the set of $(x_1, \dots, x_n)$ such that each x_i is a nonnegative integer and $\sum_{i=1}^n x_i = m$.

Remark 3. The marginal distributions have $X_i \sim \text{binomial}(m, p_i)$.

Theorem 3 (Multinomial Theorem)**.** Let m and n be positive integers and let $\mathcal{A}$ be the set of vectors $\boldsymbol{x} = (x_1, \dots, x_n)$ such that each x_i is a nonnegative integer and $\sum_{i=1}^n x_i = m$. Then for any real numbers $p_1, \dots, p_n$,

$$(p_1 + \cdots + p_n)^m = \sum_{\boldsymbol{x} \in \mathcal{A}} \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n}.$$

1.2 Exponential Families

Definition 6 (Exponential family 1)**.** A family of pmfs/pdfs is called an *exponential family* if it can be expressed

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

where $h(x) \geq 0$, the t_i are real valued functions of the observation x that do not depend on $\boldsymbol{\theta}$ and $c(\theta) \geq 0$ and the $w_i(\boldsymbol{\theta})$ are real valued functions of $\boldsymbol{\theta}$ that do not depend on x.

Theorem 4. If X is a random variable from an exponential family distribution then

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

and

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

Definition 7 (Exponential family 2)**.** We can write another parameterisation of the exponential family

$$f(x|\boldsymbol{\eta}) = h(x)c^*(\boldsymbol{\eta})\exp(\boldsymbol{\eta}\cdot\boldsymbol{t}(x))$$

where $\boldsymbol{\eta}$ is called the *natural parameter* and the set $\mathcal{H} = \{\boldsymbol{\eta} : \int_{\mathbb{R}} f(x|\eta)\mathrm{d}x < \infty\}$ is called the *natural parameter space* and is convex.

Remark 4. $\{\boldsymbol{\eta} : \boldsymbol{\eta} = \boldsymbol{w}(\boldsymbol{\theta}),\ \boldsymbol{\theta}\in\Theta\} \subseteq \mathcal{H}$. So there may be more parameterisations here than previously.

The natural parameter provides a convenient mathematical formulation, but sometimes lacks simple interpretation.

Definition 8 (Curved exponential family)**.** A *curved exponential family* distribution is one for which the dimension of $\boldsymbol{\theta}$ is $d < k$. If $d = k$ then we have a *full exponential family.*

1.3 Location and Scale Families

Definition 9 (Location family)**.** Let $f(x)$ be any pdf. The family of pdfs $f(x-\mu)$ for $\mu\in\mathbb{R}$ is called the *location family with standard pdf* $f(x)$ and μ is the *location parameter* of the family.

Definition 10 (Scale family)**.** Let $f(x)$ be any pdf. For any $\sigma > 0$ the family of pdfs $\frac{1}{\sigma}f(x/\sigma)$ is called the *scale family with standard pdf* $f(x)$ and σ is the *scale parameter* of the family.

Definition 11 (Location-Scale family)**.** Let $f(x)$ be any pdf. For $\mu\in\mathbb{R}$ and $\sigma > 0$ the family of pdfs $\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$ is called the *location-scale family with standard pdf* $f(x)$; μ is the *location parameter* and σ is the *scale parameter.*

Theorem 5 (Standardisation)**.** Let f be any pdf, $\mu\in\mathbb{R}$ and $\sigma\in\mathbb{R}_{>0}$. Then X is a random variable with pdf $\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$ if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.

Remark 5. Probabilities of location-scale families can be computed in terms of their standard variables Z

$$\mathbb{P}(X \le x) = \mathbb{P}\left(Z \le \frac{x-\mu}{\sigma}\right)$$

1.4 Inequalities and Identities

Theorem 6 (Chebychev's inequality)**.** Let X be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r > 0$,

$$\mathbb{P}(g(X)\ge r) \le \frac{\mathbb{E}[g(X)]}{r}.$$

Remark 6. This bound is conservative and almost never attained.

Remark 7 (Markov inequality)**.** The Markov inequality is the special case with $g = \mathbb{I}$.

Theorem 7. Let $X_{\alpha,\beta}$ denote a gamma(α,β) random variable with pdf $f(x|\alpha,\beta)$, where $\alpha > 1$. Then for any constants a and b:

$$\mathbb{P}(a < X_{\alpha,\beta} < b) = \beta(f(a|\alpha,\beta) - f(b|\alpha,\beta)) + \mathbb{P}(a < X_{\alpha-1,\beta} < b)$$

Lemma 2 (Stein's Lemma)**.** Let $X \sim \mathcal{N}(\theta,\sigma^2)$ and let g be a differentiable function with $\mathbb{E}[g'(x)] < \infty$. Then

$$\mathbb{E}[g(X)(X-\theta)] = \sigma^2\mathbb{E}[g'(X)]$$

The proof is just integration by parts.

Remark 8. Stein's lemma is useful for moment calculations

Theorem 8. Let χ_p^2 denote a chi squared distribution with p degrees of freedom. For any function $h(x)$,

$$\mathbb{E}[h(\chi_p^2)] = p\mathbb{E}\left[\frac{h(\chi_{p+2}^2)}{\chi_{p+2}^2}\right]$$

provided the expressions exist.

Theorem 9. Let $g(x)$ be a function that is bounded at -1 and has finite expectation, then

a. If $X \sim \text{Poisson}(\lambda)$,

$$\mathbb{E}[\lambda g(X)] = \mathbb{E}[Xg(X-1)].$$

b. If $X \sim \text{negative-binomial}(r, p)$,

$$\mathbb{E}[(1-p)g(X)] = \mathbb{E}\left[\frac{X}{r+X-1}g(X)\right].$$