1 Common Families of Distributions

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

1.1 Multinomial Distribution

Definition 1 (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 1. The marginal distributions have $X_i \sim \text{binomial}(m, p_i)$.

Theorem 1 (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 2 (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 2. 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

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

Definition 3 (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 2. $\{\boldsymbol{\eta} : \boldsymbol{\eta} = \boldsymbol{w}(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta\} \subseteq \mathcal{H}$. So there may be more parameterisations here than previously.

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

Definition 4 (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 5 (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 6 (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 7 (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 3 (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 3. Probabilities of location-scale families can be computed in terms of their standard variables Z

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

1.4 Inequalities and Identities

Theorem 4 (Chebychev's inequality)**.** Let X be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r > 0$,

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

Remark 4. This bound is conservative and almost never attained.

Remark 5 (Markov inequality)**.** The Markov inequality is the special case with $g = \mathbb{I}$.

Theorem 5. 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 1 (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 6. Stein's lemma is useful for moment calculations

Theorem 6. Let χ^2_p denote a chi squared distribution with p degrees of freedom. For any function $h(x)$,

$$\mathbb{E}[h(\chi^2_p)] = p\mathbb{E}\left[\frac{h(\chi^2_{p+2})}{\chi^2_{p+2}}\right]$$

provided the expressions exist.

Theorem 7. Let $g(x)$ be a function that is bounded at -1 and has finite expectation, then

a. If $X \sim$ Poisson(λ),

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

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

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