Families of Distributions

* We shall generally deal with families of distributions.

* A family of distributions is defined to be a set of different distributions indexed by one or more parameters.

* We can therefore study the properties of the whole family by finding the properties in terms of these parameters.

* In probability we generally specify values of the parameters but in statistics the exact values of the parameters are unknown and we use data to makes inferences about them.

Discrete Uniform Distribution

* One parameter N, a positive integer.

* $f_X(x \mid N) = \dfrac{1}{N} \qquad x = 1, \ldots, N.$

* $\mathrm{E}(X) = \dfrac{N+1}{2} \qquad \mathrm{Var}(X) = \dfrac{N^2 - 1}{12}$

* Can be transformed to any set of N consecutive integers.

Hypergeometric Distribution

- ∗ Sample K objects from N without replacement. Number of M items of interest selected.

- ∗ $f_X(x \mid N, M, K) = \dfrac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$.

- ∗ $x \in \mathbb{N} : \max\{0, K-(N-M)\} \leqslant x \leqslant \min\{K, M\}$.
  Usually $K < \min\{M, N-M\}$ so range is $0, 1, \ldots, K$.

- ∗ In that case $\mathrm{E}(X) = \dfrac{KM}{N}$ $\qquad$ $\mathrm{Var}(X) = \dfrac{KM}{N}\left(\dfrac{(N-M)(N-K)}{N(N-1)}\right)$.

Bernoulli Distribution

Definition 3.1

A Bernoulli Trial is a random experiment for which the sample space contains exactly two possible outcomes, usually labelled success and failure.

* A random variable can be defined by

$$X(\text{success}) = 1 \qquad X(\text{failure}) = 0.$$

* Such a random variable is said to have a Bernoulli(p) distribution with pmf

$$f_X(x \mid p) = p^x(1-p)^{1-x} \qquad x = 0, 1 \quad 0 \leqslant p \leqslant 1.$$

* Other random variables can be defined based on sequences of independent Bernoulli trials.

Binomial Distribution

* Suppose we run n independent Bernoulli trials each with success probability p.

* Let Y be total number of successes.

* Y is said to have a binomial(n, p) distribution with pmf

$$f_Y(y \mid n, p) = \binom{n}{y} p^y (1-p)^{n-y} \qquad y = 0, 1, \ldots, n.$$

* $\mathsf{E}(X) = np \qquad \mathsf{Var}(X) = np(1-p)$

* The moment generating function is $M_X(t) = \left[p\mathrm{e}^t + 1 - p\right]^n$.

Negative Binomial Distribution

* Run independent Bernoulli trials until observe r successes.

* Random variable is the number of trials required.

* Probability mass function

$$f_X(x \mid r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \qquad x = r, r+1, \ldots$$

* $Y = X - r$ is the number of failures before r successes

$$f_Y(y \mid r, p) = \binom{y+r-1}{r-1} p^r (1-p)^y \qquad y = 0, 1, \ldots$$

Negative Binomial Distribution

* The mean and variance are given by

$$\mathsf{E}(Y) = \frac{r(1-p)}{p} \qquad \mathsf{Var}(Y) = \frac{r(1-p)}{p^2}$$

* If we denote $\mathsf{E}(Y) = \mu$ it can be shown that

$$\mathsf{Var}(Y) = \mu + \frac{1}{r}\mu^2$$

* Often used to model overdispersion in count data.

* The geometric distribution is a special case with $r = 1$.

Poisson Distribution

* Used to model count of number of events in a time interval.

* Probability mass function

$$f_X(x \mid \lambda) = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, \ldots$$

* $E(X) = \mathrm{Var}(X) = \lambda$.

* Moment generating function

$$M_X(t) = \exp\left\{\lambda(\mathrm{e}^t - 1)\right\}$$

* Can be used to approximate the binomial distribution when $n \to \infty$, $p \to 0$, $np \to \lambda > 0$.

* Also limiting case of the negative binomial as $r \to \infty$, $p \to 1$ and $r(1-p) \to \lambda > 0$.

Uniform distribution

* Probability density function

$$f_X(x \mid a, b) = \begin{cases} \dfrac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & \text{otherwise} \end{cases}$$

* $\mathsf{E}(X) = \dfrac{b+a}{2} \qquad \mathsf{Var}(X) = \dfrac{(b-a)^2}{12}.$

* The standard uniform has $a = 0$ and $b = 1$.

Gamma Distribution

∗ The Gamma function

$$\Gamma(\alpha) = \int_0^\infty t^\alpha \mathrm{e}^{-t}\, dt \qquad \text{for } \alpha > 0$$

is a generalization of the factorial function satisfying

$$\begin{aligned} \Gamma(\alpha+1) &= \alpha\Gamma(\alpha) \\ \Gamma(n+1) &= n! \qquad \text{for any positive integer } n. \end{aligned}$$

∗ The Gamma probability density function is

$$f_X(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \mathrm{e}^{-x/\beta} \qquad 0 < x < \infty \quad \alpha > 0, \beta > 0$$

∗ $\mathrm{E}(X) = \alpha\beta \qquad \mathrm{Var}(X) = \alpha\beta^2$

∗ Moment generating function

$$M_X(t) = (1 - \beta t)^{-\alpha} \qquad t < \beta^{-1}$$

Special Cases of the Gamma Distribution

* The exponential distribution is a special case of the gamma distribution with $\alpha = 1$

$$f_X(x \mid \beta) = \frac{1}{\beta}\mathrm{e}^{-x/\beta} \qquad 0 < x < \infty$$

* The exponential random variable has the memoryless property

$$P(X > s \mid X > t) = P(X > s - t) \qquad \text{for } s > t \geqslant 0$$

* Another special case of the gamma is when $\beta = 2$ and $\alpha = p/2$ for some positive integer p. This is called the chi-squared distribution.

Normal (Gaussian) Distribution

* Probability density function

$$f_X(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad x \in \mathbb{R}$$

* If $X \sim \text{normal}(\mu, \sigma^2)$ then $Z = (X - \mu)/\sigma \sim \text{normal}(0, 1)$.

* $\text{E}(Z) = 0$ and $\text{Var}(Z) = 1$ so $\text{E}(X) = \mu$ and $\text{Var}(X) = \sigma^2$.

* The moment generating function is

$$M_X(t) = \exp\left\{\mu t + \frac{1}{2}t^2\sigma^2\right\}$$

Beta Distribution

* Two positive parameters α and β and probability density function

$$f_X(x \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad 0 < x < 1$$

* The moments of the beta distribution are

$$\mu_r' = \mathsf{E}\left(X^r\right) = \frac{\Gamma(\alpha+r)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+r)\Gamma(\alpha)}$$

* Hence we have

$$\mathsf{E}(X) = \frac{\alpha}{\alpha+\beta} \qquad \mathsf{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

* If $\alpha = \beta$ then the distribution is symmetric about $x = 0.5$.

* Taking $\alpha = \beta = 1$ gives the uniform(0,1) distribution.

Exponential Families

Definition 3.2

A family of distributions with pdf (or pmf) $f(x;\boldsymbol{\theta})$ *indexed by a vector parameter* $\boldsymbol{\theta}$ *is an* **exponential family** *distribution if we can write*

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

where $h(x) \geqslant 0$ *and* $t_1(x), \ldots, t_k(x)$ *are functions of* x *alone and* $c(\boldsymbol{\theta}) \geqslant 0$ *and* $w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ *are functions of* $\boldsymbol{\theta}$ *alone.*

The quantities $\eta_i = w_i(\boldsymbol{\theta})$ *are called the* **natural parameters** *of the family. This gives the natural parameterization*

$$f_X(x \mid \boldsymbol{\eta}) = h(x)c^*(\boldsymbol{\eta}) \exp\left\{\sum_{i=1}^{k} \eta_i t_i(x)\right\}$$

Moments of Exponential Family Distributions

Theorem 3.1

Suppose that X is a random variable from an exponential family in natural parameterization. Then

$$
\begin{aligned}
\mathsf{E}\left(t_j(X)\right) &= -\frac{\partial}{\partial \eta_j} \log c^*(\boldsymbol{\eta}) \\
\mathsf{Var}\left(t_j(X)\right) &= -\frac{\partial^2}{\partial \eta_j^2} \log c^*(\boldsymbol{\eta})
\end{aligned}
$$

Full and Curved Exponential Family Distributions

Definition 3.3

Let $\boldsymbol{\eta}$ be the d-dimensional natural parameter vector of an exponential family

$$f_X(x \mid \boldsymbol{\eta}) = h(x)c^*(\boldsymbol{\eta}) \exp\left\{\sum_{i=1}^{k} \eta_i t_i(x)\right\}$$

If $d = k$ then the family is said to be full exponential family, if $d < k$ then the family is called a curved exponential family.

* If the support $\{x : f(x \mid \boldsymbol{\theta}) > 0\}$ is a function of $\boldsymbol{\theta}$, then the family is generally not an exponential family.

* Exponential family distributions are very useful in data analysis.

Location Families

Definition 3.4

Let $f(x)$ be any probability density function and μ any real constant. Then the family of pdfs given by

$$g(x \mid \mu) = f(x - \mu)$$

is a location family with standard pdf $f(x)$ and μ is called the location parameter for the family.

* Suppose that Z has pdf $f(z)$ then the random variable $X = Z + \mu$ has pdf $g(x \mid \mu)$.

Scale Families

Definition 3.5
Let $f(x)$ be any probability density function and $\sigma > 0$ a constant. Then the family of pdfs given by

$$g(x \mid \sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$$

is a scale family with standard pdf $f(x)$ and σ is called the scale parameter for the family.

* Suppose that Z has pdf $f(z)$ then the random variable $X = \sigma Z$ has pdf $g(x \mid \sigma)$.

Location-Scale Families

Definition 3.6

Let $f(x)$ be any probability density function and $\mu \in \mathbb{R}$, $\sigma > 0$ be constants. Then the family of pdfs given by

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

is a location-scale family with standard pdf $f(x)$. μ is called the location parameter and σ is called the scale parameter for the family

* Suppose that Z has pdf $f(z)$ then the random variable $X = \mu + \sigma Z$ has pdf $g(x \mid \mu, \sigma)$.

* We cab choose the standard pdf $f(z)$ such that if Z has pdf $f(z)$ then $\mathrm{E}(Z) = 0$ and $\mathrm{Var}(Z) = 1$. In that case we have

$$\mathrm{E}(X) = \mu \qquad \mathrm{Var}(X) = \sigma^2.$$