Chapter 3: Common Families of Distributions

MATH 667-01
Statistical Inference
University of Louisville
Textbook: Statistical Inference by Casella and Berger

September 3, 2015
Last corrected: 11/19/15

- In parametric statistical models, we use families of distributions (rather than a single distribution).
- The family is indexed by one or more parameters.
- In probability, the parameter value(s) is(are) known and we can compute probabilities given a particular distribution from the family.
- In statistics, we will assume an appropriate functional form (family of distributions) and estimate the parameter value(s) to select a good distribution among the choices within the family.

- *Definition*: The *indicator function* of a set A is the function

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}.$$

Distribution	pmf $f(x)$	Mean	Variance	mgf $M_X(t)$
Uniform$\{1,\ldots,N\}$	$\frac{1}{N} I_{\{1,\ldots,N\}}(x)$	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$	$\frac{e^t(1-e^{Nt})}{N(1-e^t)}$
Binomial(n,p)	$\binom{n}{x} p^x (1-p)^{n-x} \times$ $I_{\{0,1,\ldots,N\}}(x)$	np	$np(1-p)$	$(1-p+pe^t)^n$
Hypergeometric(N,M,K)	$\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \times$ $I_{\{a,\ldots,b\}}(x)$ $a = \max\{0, M-N+K\}$ $b = \min\{K, M\}$	Kp $p = \frac{M}{N}$	$rKp(1-p)$ $r = \frac{N-K}{N-1}$	
Poisson(λ)	$\frac{e^{-\lambda}\lambda^x}{x!} I_{\{0,1,2,\ldots\}}(x)$	λ	λ	$e^{\lambda(e^t-1)}$
Negative Binomial(r,p)	$\binom{r+x-1}{x} p^r (1-p)^x \times$ $I_{\{0,1,2,\ldots\}}(x)$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{p}{1-(1-p)e^t}\right)^r$

Distribution	pmf $f(x)$	Mean	Variance	mgf $M_X(t)$
Uniform$[a, b]$	$\frac{1}{b-a} I_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{t(b-a)}$
Gamma(α,β)	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \times I_{(0,\infty)}(x)$	$\alpha\beta$	$\alpha\beta^2$	$\left(\frac{1}{1-\beta t}\right)^\alpha$
Normal(μ,σ^2)	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2	$\exp\left\{\mu t + \sigma^2 t^2/2\right\}$
Beta(α,β)	$\frac{\Gamma(\alpha+\beta) x^{\alpha-1}(1-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \times I_{(0,1)}(x)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	
Cauchy(θ)	$\frac{1}{\pi(1+(x-\theta)^2)}$	does not exist	does not exist	does not exist
Double Exponential (μ,σ)	$\frac{1}{2\sigma} e^{-\lvert x-\mu\rvert/\sigma}$	μ	$2\sigma^2$	$\frac{1}{1-t^2}$

- A family of pdfs or pmfs is called an *exponential family* if it can be expressed as

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

  Here $h(x) \geq 0$ and $t_1(x), \ldots t_k(x)$ are real-valued functions of the observation x (they cannot depend on $\boldsymbol{\theta}$), and $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ are real-valued functions of the possibly vector-valued parameter $\boldsymbol{\theta}$ (they cannot depend on x).

- *Example*: The normal distribution with mean μ and variance 1 can be expressed in the form of an exponential family. Its pdf is

$$
\begin{aligned}
f(x|\mu) &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right) e^{-\frac{1}{2}\mu^2} e^{\mu x} \\
&= h(x)c(\mu)e^{w_1(\mu)t_1(x)}
\end{aligned}
$$

  where $h(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, $c(\mu) = e^{-\frac{1}{2}\mu^2}$, $w_1(\mu) = \mu$, and $t_1(x) = x$.

- *Example*: The beta(α,β) distribution can be expressed in the form of an exponential family. Its pdf is

$$\begin{aligned}
f(x|\alpha,\beta) &= \frac{\Gamma(\alpha+\beta)x^{\alpha-1}(1-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}I_{(0,1)}(x) \\
&= I_{(0,1)}(x)\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}e^{(\alpha-1)\log x+(\beta-1)\log(1-x)} \\
&= h(x)c(\alpha,\beta)e^{w_1(\alpha,\beta)t_1(x)+w_2(\alpha,\beta)t_2(x)}
\end{aligned}$$

where $h(x) = I_{(0,1)}(x)$, $c(\alpha,\beta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$, $w_1(\alpha,\beta) = \alpha - 1$, $t_1(x) = \log x$, $w_2(\alpha,\beta) = \beta - 1$, and $t_2(x) = \log(1-x)$.

- *Example*: Consider the continuous distribution with density function

$$
\begin{aligned}
f(x|\theta) &= \frac{(\theta+1)x^{\theta}}{\theta^{\theta}}, 0 < x < \theta \\
&= \frac{(\theta+1)}{\theta^{\theta}} e^{\theta \log x}
\end{aligned}
$$

where $\theta > 0$. Is this an exponential family? Why or why not?

- *Theorem*: If X is a random variable with pdf or pmf of the form

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

then

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

and

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

- *Example*: The binomial distribution with probability of success p based on n trials can be expressed in the form of an exponential family. Here assume n is fixed. Its pmf is

$$
\begin{aligned}
f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\ldots,n\}}(x) \\
&= \left(\binom{n}{x} I_{\{0,1,\ldots,n\}}(x)\right) (1-p)^n \exp\left\{x \log\left(\frac{p}{1-p}\right)\right\} \\
&= h(x)c(p)e^{w_1(p)t_1(x)}
\end{aligned}
$$

where $h(x) = \binom{n}{x} I_{\{0,1,\ldots,n\}}(x)$, $c(p) = (1-p)^n$, $w_1(p) = \log\left(\frac{p}{1-p}\right)$, and $t_1(x) = x$.

- Alternately, the pmf can be expressed as

$$f(x|p) = \tilde{h}(x)\tilde{c}(p)e^{w_1(p)t_1(x)}$$

where $\tilde{h}(x) = \binom{n}{x}\left(\frac{1}{2}\right)^n I_{\{0,1,\dots,n\}}(x)$, $\tilde{c}(p) = 2^n(1-p)^n$, $w_1(p) = \log\left(\frac{p}{1-p}\right)$, and $t_1(x) = x$ so that $\tilde{h}(x)$ is one of the pmf's in the family.

- Directly applying the theorem to the first form, we see that

$$\mathsf{E}\left(\frac{1}{p(1-p)}X\right) = \frac{n}{1-p} \Rightarrow \mathsf{E}X = np.$$

- The set $\mathcal{H} = \left\{\eta = (\eta_1, \ldots, \eta_k) : \int_{-\infty}^{\infty} h(x) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right)\ dx < \infty\right\}$ is called the *natural parameter space* for the family. (The integral is replaced with an appropriate sum if the random variable is discrete.)
- Sometimes, an exponential family is reparametrized in terms of the natural parameter $\boldsymbol{\eta}$:

$$\begin{aligned} f(x|\boldsymbol{\eta}) &= h(x)c^*(\boldsymbol{\eta}) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right) \\ &= \frac{h(x) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right)}{\int_{-\infty}^{\infty} h(x) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right)\ dx} \end{aligned}$$

where $c^*(\boldsymbol{\eta}) = \left(\int_{-\infty}^{\infty} h(x) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right)\ dx\right)^{-1}$.

- Note that $(c^*(\boldsymbol{\eta}))^{-1}$ is the moment generating function of $(t_1(X), \ldots, t_k(X))$ if $h(x)$ is a pdf.
- Then the formulas for the first two central mometns reduce to

$$\mathsf{E}t_j(X) = -\frac{\partial}{\partial \eta_j} \log c^*(\boldsymbol{\eta})$$

and

$$\text{Var } t_j(X) = -\frac{\partial^2}{\partial \eta_j^2} \log c^*(\boldsymbol{\eta}).$$

- *Example*: In terms of the natural parameterization $\eta = \log(\frac{p}{1-p}) \Leftrightarrow p = \frac{e^\eta}{1+e^\eta}$, the pmf of the binomial distribution can be expressed as

$$f(x|\eta) = \tilde{h}(x)c^*(\eta)e^{\eta t_1(x)}$$

where $\tilde{h}(x) = \binom{n}{x}\left(\frac{1}{2}\right)^n I_{\{0,1,\ldots,n\}}(x)$, $c^*(\eta) = \left(\frac{1+e^\eta}{2}\right)^{-n}$ and $t_1(x) = x$.

- Then E$X = -\frac{\partial}{\partial\eta}\log c^*(\eta) = n\frac{e^\eta}{1+e^\eta} = np$ and Var $X = -\frac{\partial^2}{\partial\eta^2}\log c^*(\eta) = n\frac{e^\eta}{(1+e^\eta)^2} = np(1-p)$.

- *Definition*: A *full* exponential family is a family of pmf/pdf's for which the dimension of $\boldsymbol{\theta}$ is equal to k.
- *Definition*: A *curved* exponential family is a family of pmf/pdf's for which the dimension of $\boldsymbol{\theta}$ is less than k.
- *Example*: The normal family of densities with mean μ and variance σ^2 can be expressed as

$$f(x|\boldsymbol{\eta}) = h(x)c(\boldsymbol{\eta})e^{\eta_1 t_1(x)+\eta_2 t_2(x)}$$

where $h(x) = \frac{1}{\sqrt{2\pi}}$, $c^*(\boldsymbol{\eta}) = \sqrt{\eta_1}\exp\left(-\frac{\eta_2^2}{2\eta_1}\right)$, $t_1(x) = -\frac{x^2}{2}$, and $t_1(x) = x$ with $\eta_1 = 1/\sigma^2$ and $\eta_2 = \mu/\sigma^2$.

- The natural parameter space is

$$\{(\eta_1, \eta_2) : \eta_1 > 0, -\infty < \eta_2 < \infty\}$$

  so this is a full exponential family.

- If we assume the $\mu = \sigma$, then we obtain a one-dimensional curved exponential family

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\mu^2}} e^{-1/2} \exp\left(-\frac{x^2}{2\mu^2} + \frac{x}{\mu}\right)$$

  with parameter space

$$\left\{(\mu, \mu^2) : -\infty < \mu < \infty\right\}.$$

- *Theorem*: Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constants. Then the function

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

is a pdf.

- *Definition*: Let $f(x)$ be any pdf. Then the family of pdfs $f(x-\mu)$, indexed by the parameter μ, $-\infty < \mu < \infty$, is called the *location family with standard pdf* $f(x)$ and μ is called the *location parameter* for the family.
- *Definition*: Let $f(x)$ be any pdf. Then for any $\sigma > 0$, the family of pdfs $(1/\sigma)f(x/\sigma)$, indexed by the parameter σ, is called the *scale family with standard pdf* $f(x)$ and σ is called the *scale parameter* of the family.
- *Definition*: Let $f(x)$ be any pdf. Then for any μ, $-\infty < \mu < \infty$, and any $\sigma > 0$, then family of pdfs $(1/\sigma)f((x-\mu)/\sigma)$, indexed by the parameter (μ, σ), is called the *location-scale family with standard pdf* $f(x)$; μ is called the *location parameter* and σ is called the *scale parameter*.

- *Theorem*: Let $f(\cdot)$ be any pdf. Let μ be any real number and let σ be any positive real number. Then X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$ if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.
- *Theorem*: Let Z be a random variable with pdf $f(z)$. Suppose $\mathrm{E}Z$ and Var Z exist. If X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$, then

$$\mathrm{E}X = \sigma \mathrm{E}Z + \mu \text{ and } \mathrm{Var}\ X = \sigma^2 \mathrm{Var}\ Z.$$

- *Example*: The gamma distribution with a location parameter μ has density

$$\frac{1}{\Gamma(\alpha)\beta}\left(\frac{x-\mu}{\beta}\right)^{\alpha-1} e^{-(x-\mu)/\beta} I_{(\mu,\infty)}.$$

This is a location-scale family with mean

$$\mathsf{E}X = \mu + \frac{\alpha}{\beta}$$

and

$$\mathsf{Var}\ X = \frac{\alpha}{\beta^2}$$

- *Theorem*: Chebychev's Inequality: Let X be a random variable such that $\mathsf{E}g(X)$ exists and let $g(x)$ be a nonnegative function. Then for any $r > 0$,

$$P(g(X) \geq r) \leq \frac{\mathsf{E}g(X)}{r}.$$

- Here are two common special cases of Chebychev's inequality:
  1. Markov's Inequality: $P(|X| \geq r) \leq \dfrac{\mathsf{E}|X|}{r}$
  2. $P(|X - \mu| \geq t\sigma) \leq \dfrac{1}{t^2}$.