Special Distributions

Brian Vegetabile

2016 Statistics Bootcamp
Department of Statistics
University of California, Irvine

September 7th, 2016

Bernoulli Distribution I

- One of the most important distributions in statistics is the Bernoulli Distribution
- The Bernoulli distribution is used to describe experiments with binary outcomes, say 0 and 1.
  - Think 'heads' or 'tails', 'yes' or 'no', 'win' or 'loss'
  - Often called a 'Bernoulli trial'
- Ultimately, there is some probability p of 'succeeding' and a corresponding probability $(1-p)$ of failing based upon the rules of probability.

Bernoulli Distribution II

- If we define the value 1 as being a success, we can write this as follows

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}, \quad 0 \leq p \leq 1$$

- To create a probability mass function, consider

$$P[X = 1] = p \qquad P[X = 0] = 1 - p$$

therefore one way to write the mass function is as follows

$$P[X = x] = p_X(x) = \begin{cases} p^x(1-p)^{1-x} & x = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

- *Show properties of this distribution: CDF, expectation, variance, MGF...*

Bernoulli Distribution III

- It is easy to see that this is a probability mass function.
  - $p_X(x) \geq 0$ for all x, and
  - $\sum_x p_X(X) = p + (1-p) = 1$.
- We can also easily find the mean and variance,

$$\begin{aligned} E(X) &= \sum_x x p_X(x) = 1 \times (p) + 0 \times (1-p) = p \\ E(X^2) &= \sum_x x^2 p_X(x) = 1^2 \times (p) + 0 \times (1-p) = p \\ Var(X) &= E(X^2) - (E(X))^2 = p - p^2 = p(1-p) \end{aligned}$$

- Additionally, we can find the moment generating function for this random variable

$$E(e^{tX}) = \sum_x e^{tx} p_X(x) = e^{t(1)} p + e^{t(0)} (1-p) = (1-p) + pe^t$$

Binomial Distribution I

- Related to the Bernoulli distribution is the Binomial Distribution.
- A binomial random variable can arise from a sequence of Bernoulli trials with the properties that,
  - Trials are independent events
  - Each trial results in exactly one of the same two mutually exclusive outcomes
  - The probability of success (and subsequently failure) remains constant from trial to trial.
- Therefore a binomial random variable can be considered as the sum of n Bernoulli random variables. That is the number of successes in n Bernoulli trials.
  - Example: Number of 'heads' in ten independent coin tosses.

Binomial Distribution II

- We can write the probability mass function in a similar way to the Bernoulli distribution

$$P[X=x] = p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0,1,2,\ldots,n \\ 0 & \text{otherwise} \end{cases}$$

- Note: Showing that this is indeed a distribution requires the use of the binomial theorem, where

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$$

- The expectation and variance are also similar

$$E(X) = np \qquad Var(X) = np(1-p)$$

Poisson Distribution I

- Another important discrete distribution is the Poisson distribution.
- While the Binomial distribution counts the number of successes in a series of trials, the Poisson distribution counts the number of events in a given time interval.
  - Binomial ‘counts’ are bounded by the number of trials
  - Poisson counts are in an interval are not bounded.
- Examples that generally can be modeled with a Poisson Distribution
  - The number of misprints on a page (or a group of pages) of a book
  - The number of customers entering a post office on a given day
  - The number of α-particles discharged in a fixed period of time from some radioactive material

Poisson Distribution II

- Additionally, the Poisson distribution can be used to model the number of events that occur in a spatial region.
- The distribution is parameterized by a value λ which is often referred to as the rate or intensity of the distribution, which governs the mean of the distribution
- The mass function is given as follows

$$f(x|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \text{for } x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Poisson Distribution III

- To verify that this is a distribution, we must show that $\sum_{x=0}^{\infty} f(x|\lambda) = 1$. Additionally, from calculus, we know the power series characterization $e^a = \sum_{n=0}^{\infty} \frac{a^n}{n!}$. Thus,

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda \lambda^x}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

- We can use similar mathematical tricks to derive the mean and variance.
- The Poisson distribution can be used to approximate the Binomial distribution.

Self-Study: Review Poisson Process

- The Poisson distribution can be derived from a few basic assumptions that we will outline below.
- In the interest of time we will only list the basic assumptions:
  i) Start with no arrivals
  ii) Arrivals in disjoint time periods are independent
  iii) Number of arrivals depends only on the period length
  iv) Arrival probability is proportional to the period length, if length is small
  v) No simultaneous arrivals

Uniform Distribution

- The simplest continuous distribution is when mass is spread out 'uniformly' on some interval $[a, b]$
- The density function is as follows:

$$f(x|\lambda) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

- Quickly show CDF and Expected Values

Normal Distribution I

- The most famous and often used distribution would be the Normal distribution and informally can be referred to as the 'bell curve'
- The distribution is symmetric and unbounded on the real line, but concentrates mass at it's mean/mode/median.
- It is very useful and can be used to satisfactorily represent many phenomenon in the world such as
  - Distribution of heights of Airforce Pilots
  - Distribution of IQ scores
  - Distribution of measurement errors
- Finally the distribution plays an important role in the central limit theorem which is used in much of statistics.

Normal Distribution II

- The density of the distribution is

$$f(x|\lambda) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \text{ for } -\infty < x < \infty$$

- The following are the mean and variance of the distribution

$$E(X) = \mu \qquad Var(X) = \sigma^2$$

- $\sqrt{\sigma^2} = \sigma$ is often referred to the as standard deviation of the distribution.
- We do not derive these properties here.

Gamma Distribution I

- The Gamma distribution is important for a number of reasons.
- First it is a positive valued function that can be used to model phenomenon.
- Additionally, specific parameter values of the Gamma distribution create many other named distributions which are also important. (exponential, Weibull, χ^2, etc)
- The Gamma distribution allows plays important roles throughout Bayesian Statistics.

Gamma Distribution II

- An important relationship for this distribution is that of the gamma function, specifically provided α is positive,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

which satisfies

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$$

and finally for any integer $n > 1$, $\Gamma(n) = (n-1)!$.

Gamma Distribution III

- The density of the family of gamma distributions is as follows

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)$$

  where α is the shape parameter since it controls the 'peakedness' of the distribution and β is called the scale since it mainly influences the spread of the distribution.

Kernel Trick for Integration I

- To illustrate the 'kernel trick' for integration, we find the expected value of the gamma distribution.

$$
\begin{aligned}
E(X) &= \int_0^\infty x x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{(\alpha+1)-1} \exp\left(-\frac{x}{\beta}\right) dx
\end{aligned}
$$

- We notice though that if we multiply and divide by $\frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}}$, then the integral becomes the pdf of a Gamma$(\alpha+1, \beta)$ distribution.

$$
= \frac{\Gamma(\alpha+1)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^{(\alpha+1)-1} \exp\left(-\frac{x}{\beta}\right) dx
$$

Kernel Trick for Integration II

- The term on the right integrates to 1 and we are left with the following expression.

$$
\begin{aligned}
&= \frac{\Gamma(\alpha+1)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^{\alpha}} \\
&= \alpha\beta
\end{aligned}
$$

  Where the last line holds by properties of the gamma function.

- The kernel trick will become invaluable through the course of the year.

Special Gamma Distributions

- Within the gamma(α, β) family, there are many named special distributions.
- When $\alpha = 1$, the gamma distribution reduces to the exponential distribution
- If $\alpha = p/2$, where p is an integer, and $\beta = 2$, then the gamma distribution becomes a χ^2 distribution with p degrees of freedom
  - The χ^2 distribution will become very important throughout the year.
- The list goes on and on....

Beta Distribution I

- Another important distribution that will come up often is the Beta distribution which a continuous and bounded random variable.
- The density is continuous on the interval $(0, 1)$ and is indexed by the parameters α and β.
- Most frequently used in Bayesian statistics to model a priori beliefs about proportions.
- There is a more general family of beta distributions for general intervals

Beta Distribution II

- The distribution relies on the relationship

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx.$$

  where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.
- Thus the density is

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)}x^{\alpha-1}(1-x)^{\beta-1} \text{ for } x \in [0, 1], \alpha > 0, \beta > 0.$$

- When $\beta = \alpha = 1$ the beta reduces to the Uniform distribution on $(0, 1)$.

Bivariate Normal Distributions

- To introduce multivariate distributions, we define the bivariate normal distribution.
- A RV $\mathbf{X} = (X_1, X_2)$ has the bivariate normal distribution $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ if (for some $\sigma_i > 0, -1, \rho < 1)$ and real-valued μ_i

$$
\begin{aligned}
f(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) \quad = \quad & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \right.\right. \\
& \left.\left. 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right)
\end{aligned}
$$

- When $\rho = 0$ this will factor into two independent normal distributions.

References