Rice University Summer Institute of Statistics
RUSIS Lab 5

Statistical Computing Sessions

June 1, 2011

1 Remark on Notation: Normal Distribution

If X is a random variable with mean μ and variance σ^2, written $X \sim \mathcal{N}(\mu, \sigma^2)$, the probability density function (pdf) of X is

$$f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma}(x-\mu)^2}, \qquad -\infty < x < \infty$$

The notation $f_X(x|\mu, \sigma^2)$ is derived from the following conventions. The lower-csae f denotes the fact that we are referring to a pdf (or a pmf), the subscript $_X$ specifies that the pdf is of the random variable X. Inside the argument of the function, there is a vertical bar | (sometimes a colon : or semicolon ; is used instead) which distinguishes the independent variables (here x) from the parameters (here μ and σ) which are considered to be known. This is sometimes also written $f_{\mu,\sigma^2}(x)$ if it is understood that f_{μ,σ^2} is the pdf of X.

2 Beta Distribution

The beta distribution is a continuous distribution with support $[0, 1]$ and thus is useful for modeling uncertainty in that range. The beta distribution has two parameters, α and β, which are both considered shape parameters (and the exercises will show their effects).

Any combination of $\alpha > 0$ and $\beta > 0$ provides a valid probability distribution. If X is a random variable with shape parameters α and β, written $X \sim Beta(\alpha, \beta)$, X exhibits the probability density function (pdf)

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

where $\Gamma(x)$ is the "Gamma function,"

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

The gamma function is used throughout advanced mathematics and exhibits many interesting properties. For example, the gamma function is a generalization of the factorial "function" for positive integers since

$$\Gamma(n) = (n-1)!$$

and perhaps this give insight into why $0! = 1$, since

$$0! = (1-1)! = \Gamma(1) = \int_0^\infty t^{1-1} e^{-t} dt = \int_0^\infty e^{-t} dt = 1$$

In statistics, the beta distribution is particularly useful in Bayesian statistics, which will be discussed later in the lectures (or maybe not!!). For now, it suffices to say that it has interesting and fruitful connections with the binomial distribution which we have already discussed. A bit more specifically, the beta distribution provides a natural conjugate prior for the parameter p; but let's not get ahead of ourselves.

1. One of the beta distributions is another contender for the most useful distribution in statistics. This is the Beta(1,1) distribution, but we usually call it a different name. Input the pdf of the beta distribution using Mathematica command `f[x_, a_, b_] :=...`, and use the `Plot` command to plot this special case of the beta distribution. What do we usually call it? (What is the Beta(1,1)?)

2. Use the `Manipulate` and `Plot` commands, fix β at 1 and animate the pdf as α ranges from .1 to 10. Be sure to fix `PlotRange -> {0, 5}` within the `Plot` command.

3. Now fix α at 1 and animate the pdf as β ranges from .1 to 10.

4. Animate the pdf as both α and β range from .1 to 10.

The beta density can take so many different shapes depending on the values of α and β. Check the following combinations and see what shape the pdf of Beta takes.

- (Again) $\alpha = 1$ and $\beta = 1$
- $\alpha < 1, \beta < 1$
- $\alpha < 1, \beta \geq 1$ or $\alpha = 1, \beta > 1$
- $\alpha = 1, \beta < 1$ or $\alpha > 1, \beta \leq 1$
- $\alpha > 1, \beta > 1$

3 Fundamental Theorem of Simulation

One very important theorem in statistics is the so-called Fundamental Theorem of Simulation (FTS). It's statement and proof (for the simple univariate case) are as follows.

> THM. (FUNDAMENTAL THEOREM OF SIMULATION) Let $X \sim F_X(x)$ and $F_X(x)$ be a monotone function and $U \sim Unif(0,1)$. Then $F_X^{-1}(U) \sim F_X$.

Proof : Suppose that $F_X(x)$ is a monotone increasing function (monotone decreasing is proved similarly.) Then by the definition of the cdf,

$$\begin{aligned} F_{F_X^{-1}}(x) &= P[F_X^{-1}(U) \leq x] \\ &= P[U \leq F_X(x)] \\ &= F_U(F_X(x)) \\ &= F_X(x) \end{aligned}$$

since $0 \leq F_X(x) \leq 1$ and $F_U(u) = u$ in this interval.

The result is also sometimes called the "Probability Integral Transform" - but what does it mean? Throughout statistics, it is often very useful to simulate random numbers from various distributions. However, it it not very clear how one might generate random samples from a particular distribution. In many cases, the

FTS gives us a way to do just that provided we can generate $Unif(0,1)$ samples (which we will just assume we can do). Let's consider the example of the exponential distribution.

Suppose you want to generate 100 samples from the distribution $Exp(\lambda)$, but you only know how to generate samples from the $Unif(0,1)$ distribution. From the discussion in the section on the gamma distribution, we know that the pdf of the (rate-parameterized) exponential distribution is

$$f_X(x|\lambda) = f_X(x|1,\lambda) = \frac{\lambda^1}{\Gamma(1)} x^{1-1} e^{\lambda x} = \lambda e^{-\lambda x}, \qquad x > 0.$$

To find the cdf $F_X(x)$, we simply integrate the pdf from $-\infty$ to x (careful to integrate the correct pdf - the pdf given above is 0 when x is less than zero!). Thus, if $x \leq 0, F_X(x) = 0$, and for $x > 0$,

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t|\lambda) dt \\ &= \int_0^x \lambda e^{-\lambda t} dt \\ &= 1 - e^{-\lambda x}, \qquad x > 0. \end{aligned}$$

But, for the FTS, what we really want is $F_X^{-1}(x)$, not $F_X(x)$, so we simply do the swapping the x and y technique and solve for y:

$$x = 1 - e^{-\lambda y} \Rightarrow e^{-\lambda y} = 1 - x \Rightarrow -\lambda y = log(1-x) \Rightarrow y = -\frac{-1}{\lambda} log(1-x)$$

Thus,

$$F_X^{-1}(x) = \frac{-1}{\lambda} log(1-x), \qquad 0 < x < 1.$$

Now, by the FTS, if we take a sample u from the $Unif(0,1)$ distribution (note that samples are generally written with lower case letters, and random variables with upper case letters), and apply F_X^{-1} to it, we will have a sample from the exponential distribution with rate parameter λ.

i. Using either the `rbeta` command or the `runif` command, simulate 100 samples from the uniform distribution on the unit interval, i.e., from the distribution $Unif(0,1)$. In lieu of a histogram, make a density estimate from the samples (that is, a smoothed histogram), using `geom = 'density'`. (As in `qplot(samples, geom = 'density')`.)

ii. Use these 100 uniformly distributed samples to generate 100 exponentially distributed samples with rate parameter $\lambda = 2$. Make a density estimate of the sample as in the previous example. Since we know the true underlying pdf, we can add it to the plot and see how well the density estimate approximates the true density. Using the `dexp` and `stat_function` commands, add the true density curve to the histogram in red. How good is the approximation?

iii. Repeat the above exercise with n = 1000 and n = 10000 samples. How does the approximation change?

iv. Finally, repeat the above with histograms instead of density estimates. Which method do you prefer for understanding the underlying distribution?