

STA 623 – Fall 2009 – Dr. Charnigo

Section 3.3: Continuous Distributions

Normal distribution. A random variable X has the normal distribution with mean $\mu \in (-\infty, \infty)$ and standard deviation $\sigma \in (0, \infty)$ if its probability density function is

$$f(x) = (2\pi)^{-1/2}\sigma^{-1} \exp[-(x - \mu)^2/(2\sigma^2)].$$

Suppose that X has the normal distribution with mean μ and standard deviation σ . Let $a \in (-\infty, 0) \cup (0, \infty)$ and $b \in (-\infty, \infty)$. Then we can readily verify that $aX + b$ has the normal distribution with mean $a\mu + b$ and standard deviation $|a|\sigma$. A special case is $a := \sigma^{-1}$ and $b := -\mu\sigma^{-1}$, which yields the normal distribution with mean 0 and standard deviation 1 (called the standard normal distribution).

The practical importance of the above result is that any probability involving a normal random variable can be expressed as a probability involving a standard normal random variable, and the cumulative distribution function of a standard normal random variable has been tabulated. An abbreviated table is shown below, where Z denotes a standard normal random variable. The process of defining $Z := \sigma^{-1}X - \mu\sigma^{-1} = (X - \mu)/\sigma$ and calculating a probability involving X in terms of Z is called standardization.

Table 1:

z	$P(Z \leq z)$	z	$P(Z \leq z)$
-3	0.0013	1	0.8413
-2	0.0228	2	0.9772
-1	0.1587	3	0.9987
0	0.5000		

To illustrate, suppose that X has the normal distribution with mean 100 and standard deviation 10. What is $P(X \geq 70)$? What is $P(90 \leq X \leq 120)$?

Normal distributions are of special interest for two reasons. First, many physical, biological, or social phenomena can reasonably be modeled using a normal distribution. Second, if a random variable X can be expressed as a sum of independent random variables X_1, \dots, X_n , then under fairly general conditions the cumulative distribution function of X can be approximated by the cumulative distribution function of a normal random variable with mean $E[X]$ and standard deviation $\sqrt{Var[X]}$. This is a consequence of the Central Limit Theorem, with which you will become familiar in STA 606.

For instance, we know that a binomial random variable X with parameters p and n can be expressed as $X_1 + \dots + X_n$, where $X_i := 1_{\{\text{success on trial } i\}}$ for $i \in \{1, \dots, n\}$. Since X has mean np and standard deviation $\sqrt{np(1-p)}$, the Central Limit Theorem tells us that the cumulative distribution function of X can be approximated by the cumulative distribution function of a normal random variable with mean np and standard deviation $\sqrt{np(1-p)}$. Or, put differently, $(X - np)/\sqrt{np(1-p)}$ “looks” like a standard normal random variable. The quality of the approximation gets better as $np(1-p)$ gets larger.

To illustrate, suppose that X has the binomial distribution with parameters $p = 0.5$ and $n = 100$. Then $np = 50$ and $\sqrt{np(1-p)} = \sqrt{25} = 5$. Letting Z denote a standard normal random variable, we have

$$P(45 \leq X \leq 55) = P(-1 \leq (X - 50)/5 \leq 1) \approx P(-1 \leq Z \leq 1) =$$

Actually, since $P(45 \leq X \leq 55) = P(45 - \delta < X < 55 + \delta)$ for any $\delta \in (0, 1]$, we can validly approximate this probability by $P(-1 - \delta/5 < Z < 1 + \delta/5)$. Such a δ is referred to as a continuity correction. The best choice of δ is arguably 0.5, on the grounds that $5Z + 50$ is meant to “look” like X , so $X = 45$ should translate to $44.5 < 5Z + 50 < 45.5$ rather than to (say) $5Z + 50 = 45$ or $44 < 5Z + 50 < 46$. In fact, with $\delta = 0.5$, we obtain $P(-1.1 < Z < 1.1) = 0.7287$, which is in agreement with the actual value of $P(45 \leq X \leq 55)$ to four decimal places.

Gamma distribution. A random variable X has the gamma distribution with parameters $\alpha \in (0, \infty)$ and $\beta \in (0, \infty)$ if its probability density function is

$$f(x) = \frac{1}{\Gamma[\alpha]\beta^\alpha} x^{\alpha-1} \exp[-x/\beta] 1_{\{x>0\}}.$$

We have $E[X] = \alpha\beta$ and $Var[X] = \alpha\beta^2$. We refer to α as a shape parameter and to β as a scale parameter.

An alternative parametrization replaces β with $1/\lambda$, where $\lambda \in (0, \infty)$. Then $E[X] = \alpha/\lambda$ and $Var[X] = \alpha/\lambda^2$. When necessary to distinguish between the two parametrizations, we call the one with β a “mean” parametrization (because $E[X]$ is proportional to β) and the one with λ a “rate” parametrization (for reasons that will emerge in STA 624, if you take that course).

Worth noting is that the shape of $f(x)$ is highly sensitive to α , which is why α is called a shape parameter. When $\alpha \in (0, 1]$, $f(x)$ is strictly decreasing on $(0, \infty)$. When α exceeds 1, $f(x)$ has a mode — i.e., a point at which $f(x)$ is maximized — interior to $(0, \infty)$. As α continues increasing, the mode of $f(x)$ moves rightward and $f(x)$ takes on a bell-shaped appearance. In fact, for very large α , a gamma distribution is well approximated by a normal distribution. Thus, for purposes of modeling physical, biological, or social phenomena not well described by a normal distribution, a gamma distribution with a small or modest α is a more viable choice than a gamma distribution with a large α .

In the special case that $\alpha = 1$, we say that X has the exponential distribution with scale parameter β .

In the special case that $\alpha = p/2$ and $\beta = 2$, where p is a positive integer, we say that X has the chi-square distribution on p degrees of freedom. (In fact, other than for convenience in producing tables for the backs of methods textbooks, there is no real reason that a chi-square distribution must have integer degrees of freedom. So, if we like, we can just let p be a positive real.) What are the mean and standard deviation of a chi-square random variable on p degrees of freedom? What will X/p “look” like when p is large?

Weibull distribution. Let $\gamma \in (0, \infty)$. If X has the exponential distribution with scale parameter $\beta \in (0, \infty)$, then $Y := X^{1/\gamma}$ has probability density function

$$f(y) = \frac{\gamma}{\beta} y^{\gamma-1} \exp[-y^\gamma/\beta] 1_{\{y>0\}}$$

and is said to have the Weibull distribution with parameters γ and β . We have $E[Y] = \beta^{1/\gamma} \Gamma[1 + 1/\gamma]$ and $Var[Y] = \beta^{2/\gamma} \{\Gamma[1 + 2/\gamma] - \Gamma^2[1 + 1/\gamma]\}$.

For any positive real y we have

$$P(Y \leq y) = \int_0^y \frac{\gamma}{\beta} t^{\gamma-1} \exp[-t^\gamma/\beta] dt =$$

We then have

$$S(y) := P(Y > y) =$$

and

$$H(y) := -\frac{d}{dy} \log S(y) =$$

We refer to $S(y)$ as a survival function and to $H(y)$ as a hazard function. We often interpret Y as the lifetime of a person or object, although Y can also be interpreted as the time until some generic event of interest occurs. Retaining the former interpretation for now, the survival function returns the probability that a person or object lives more than y time units. To understand the hazard function, let δ be a small positive number and consider $P(Y \leq y + \delta | Y > y)$. In words, this is the probability of a person or object dying in the next δ time units given that the person or object is alive at time y . We have

$$P(Y \leq y + \delta | Y > y) = \frac{P(y < Y \leq y + \delta)}{P(Y > y)} \approx \frac{\delta f(y)}{S(y)} = \delta H(y).$$

Hence, a hazard function that increases over $(0, \infty)$ signifies a belief that an older person or object is more likely to expire over a short time interval than a younger person or object.

Given the above calculations, what can we say about the potential applicability of a Weibull distribution to describe a lifetime?

Beta distribution. A random variable X has the beta distribution with parameters $\alpha \in (0, \infty)$ and $\beta \in (0, \infty)$ if its probability density function is

$$f(x) = \frac{\Gamma[\alpha + \beta]}{\Gamma[\alpha]\Gamma[\beta]} x^{\alpha-1} (1-x)^{\beta-1} 1_{\{0 < x < 1\}}.$$

We have $E[X] = \alpha/(\alpha + \beta)$ and $Var[X] = \alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$.

In the special case that $\alpha = \beta = 1$, we say that X has the uniform distribution on $(0, 1)$. Moreover, if a and b are any reals with $a < b$, then $a + (b - a)X$ is said to have the uniform distribution on (a, b) .

A beta distribution is sometimes employed in microarray data analysis. For instance, suppose that there are 5000 genes and that, for each gene, we have tested a null hypothesis that the mean expression level of the gene is the same within each of two populations, say people with and without a particular illness. (Here I am taking for granted that you already have some knowledge of hypothesis testing from STA 602. You will study hypothesis testing from a more theoretical perspective in STA 606, but that theoretical perspective is not required here.) Thus, for each gene we have obtained a p-value. If all 5000 null hypotheses are true, then the p-values should be uniformly distributed on $(0, 1)$. However, if some of the null hypotheses are false, then there should be more small p-values than large p-values. What can we say about the potential applicability of a beta distribution to describe such a collection of p-values?

Cauchy distribution. A random variable X has the Cauchy distribution with parameters $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$ if its probability density function is

$$f(x) = \frac{1}{\pi\sigma} \frac{1}{1 + (x - \mu)^2/\sigma^2}.$$

We cannot call μ a mean since $E[X]$ does not exist as a finite number, nor can we call σ a standard deviation since $E[X^2]$ does not exist as a finite number. Instead, we call μ a location parameter and σ a scale parameter. Actually, we can be more specific and call μ a median, since $P(X \leq \mu) = P(X \geq \mu) = 1/2$.

Lognormal distribution. A random variable X has the lognormal distribution with parameters $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$ if its probability density function is

$$f(x) = (2\pi)^{-1/2} \sigma^{-1} x^{-1} \exp[-(\log x - \mu)^2 / (2\sigma^2)] 1_{\{x>0\}}.$$

The lognormal distribution is so named because $\log X =: Y$ has the normal distribution with mean μ and standard deviation σ . Thus, lognormal distributions are appealing models for physical, biological, or social phenomena whose quantifications are strictly positive (so that their logarithms are defined) with logarithms that are well described by a normal distribution.

Recalling that $M_Y(t) = \exp[\mu t + \sigma^2 t^2 / 2]$, we see that

$$E[X] = E[\exp Y] = \exp[\mu + \sigma^2 / 2] \quad \text{and}$$

$$\begin{aligned} \text{Var}[X] &= E[X^2] - E[X]^2 \\ &= \end{aligned}$$

Double exponential distribution. A random variable X has the double exponential distribution with parameters $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$ if its probability density function is

$$f(x) = (2\sigma)^{-1} \exp[-|x - \mu| / \sigma].$$

While unimodal (i.e., possessing only one mode), $f(x)$ is not bell-shaped.

Putting $y := x - \mu$, we have

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x (2\sigma)^{-1} \exp[-|x - \mu| / \sigma] dx \\ &= \int_{-\infty}^{\infty} (y + \mu) (2\sigma)^{-1} \exp[-|y| / \sigma] dy \\ &= \end{aligned}$$

We can also show that $\text{Var}[X] = 2\sigma^2$.