Math 425
Introduction to Probability
Lecture 21

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

March 6, 2009

Functions of a Random Variable: Expectation

☞ There are an overwhelming number of possible random variables which can be derived from even a single distribution, such as U, a uniformly distributed r.v., by applying a function, $g(U)$.

It is nice to know that we need only one theorem to dispose of expectation:

Theorem

Let X be a continuous random variable with density f_X and $g : \mathbb{R} \to \mathbb{R}$ a continuous function (except, perhaps on finitely many points). Then

$$E[g(X)] = \int_{-\infty}^{\infty} g(t) f_X(t)\, dt$$

provided this integral exists.

Functions of a Random Variable: Cumulative Distribution

☞ Suppose we would like to compute the cumulative distribution of a random variable $g(X)$, when we have F_X.

One theorem is all we need to know.

Theorem

Let X be a continuous random variable, and $g(t)$ a strictly increasing function on the range of X. Let $Y = g(X)$.

The cumulative distribution of Y is

$$F_Y(t) = F_X(g^{-1}(t)).$$

If $g(x)$ is strictly decreasing on the range of X, then

$$F_Y(t) = 1 - F_X(g^{-1}(t)).$$

Proof

Suppose $g(t)$ is strictly increasing. Then g is 1-1, so has an inverse (and both will be differentiable on all but a discrete set of points).

$$\begin{aligned} F_Y(t) &= \mathbf{P}\{Y \le t\} \\ &= \mathbf{P}\{g(X) \le t\} \\ &= \mathbf{P}\{X \le g^{-1}(t)\} \\ &= F_X(g^{-1}(t)) \end{aligned}$$

Suppose $g(t)$ is strictly decreasing.

$$\begin{aligned} F_Y(t) &= \mathbf{P}\{Y \le t\} \\ &= \mathbf{P}\{g(X) \le t\} \\ &= \mathbf{P}\{g^{-1}(t) \le X\} \\ &= 1 - \mathbf{P}\{X \le g^{-1}(t)\} \\ &= 1 - F_X(g^{-1}(t)) \end{aligned}$$

Functions of a Random Variable: Density

☞ A function $g(t)$ is monotonic on its domain if it is either strictly increasing or strictly decreasing on its domain.

Theorem

Let X be a continuous random variable, and $g(t)$ a monotonic function function on the range of X. Let $Y = g(X)$.
The density of Y is

$$f_Y(t) = f_X(g^{-1}(t)) \cdot \left|\frac{d}{dt}g^{-1}(t)\right|.$$

Note: Compare to Ross, Theorem 5.7.1, page 243.

Proof

Suppose $g(t)$ is strictly increasing (so $\frac{d}{dt}g^{-1}(t) > 0$):

$$\begin{aligned} f_Y(t) &= \frac{d}{dt}F_Y(t) \\ &= \frac{d}{dt}F_X(g^{-1}(t)) \\ &= f_X(g^{-1}(t)) \cdot \left|\frac{d}{dt}g^{-1}(t)\right| \qquad \text{(chain rule)} \end{aligned}$$

Suppose $g(t)$ is strictly decreasing (so $\frac{d}{dt}g^{-1}(t) < 0$):

$$\begin{aligned} f_Y(t) &= \frac{d}{dt}F_Y(t) \\ &= \frac{d}{dt}(1 - F_X(g^{-1}(t))) \\ &= f_X(g^{-1}(t)) \cdot \left|\frac{d}{dt}g^{-1}(t)\right| \qquad \text{(chain rule)} \end{aligned}$$

Example: Uniform Distribution

Example. There is a uniform distribution for each interval $[\alpha, \beta]$.
The only one you need to know is the uniform distribution U on $[0, 1]$.

☞ Let V be the uniformly distributed r.v. on $[\alpha, \beta]$.
Define $g(t)$ by

$$g(t) = (\beta - \alpha)t + \alpha \quad \text{so,} \quad g^{-1}(t) = \frac{t - \alpha}{\beta - \alpha}.$$

Then $V = g(U)$:
$t \in [0, 1] \quad \mapsto \quad (\beta - \alpha)t + \alpha \in [\alpha, \beta]$

1. Scale by $\beta - \alpha$ to $(\beta - \alpha)t$.
2. Shift by α to $(\beta - \alpha)t + \alpha$.

Example: Uniform Distribution

$$g(t) = (\beta - \alpha)t + \alpha \qquad g^{-1}(t) = \frac{t - \alpha}{\beta - \alpha} \qquad \frac{d}{dt}g^{-1}(t) = \frac{1}{\beta - \alpha}$$

☞ U is uniformly distributed on $[0, 1]$.

$$F_U(t) = t \qquad f_U(t) = 1$$
$$E[V] = \frac{1}{2} \qquad Var(V) = \frac{1}{12}.$$

☞ $V = g(U)$ is uniformly distributed on $[\alpha, \beta]$.

$$F_V(t) = F_U(g^{-1}(t)) = \frac{t - \alpha}{\beta - \alpha} \qquad f_V(t) = \frac{1}{\beta - \alpha}$$
$$E[V] = E[g(t)] = \frac{(\beta - \alpha)}{2} + \alpha \qquad Var(V) = \frac{(\beta - \alpha)^2}{12}.$$

☞ Even if g is not strictly increasing or decreasing, we can often apply the same technique (although the situation may be more complicated).

Example. Let X be a continuous r.v. with known cumulative distribution F_X and density f_X. Let $Y = X^2$.

$$\begin{aligned} F_Y(t) &= \mathbf{P}\{Y \le t\} \\ &= \mathbf{P}\{-\sqrt{t} \le X \le \sqrt{t}\} \\ &= \mathbf{P}\{X \le \sqrt{t}\} - \mathbf{P}\{X \le -\sqrt{t}\} \\ &= F_X(\sqrt{t}) - F_X(-\sqrt{t}) \end{aligned}$$

$$\begin{aligned} f_y(t) &= \frac{d}{dt} F_y(t) \\ &= \frac{d}{dt}\left[F_X(\sqrt{t}) - F_X(-\sqrt{t})\right] \\ &= \left[f_x(\sqrt{t}) + f_x(-\sqrt{t})\right] \cdot \frac{1}{2\sqrt{t}} \end{aligned}$$

Normal Density

☞ The most important density function is the normal density function.

Definition

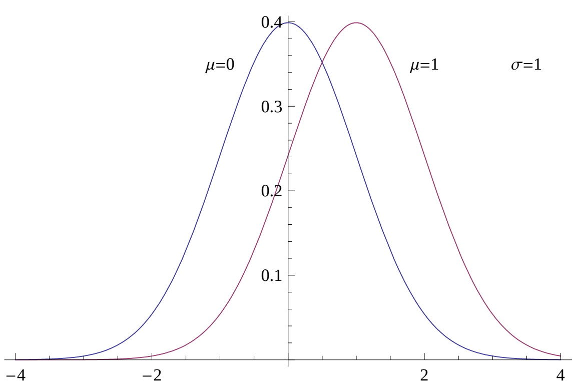
The normal density function with parameters μ and σ^2 is defined as

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/2\sigma^2} \qquad \text{for every } t \in \mathbb{R}.$$

A random variable is said to be normally distributed with parameters μ and σ^2, if its density function is a normal density function for some parameters μ and σ^2.

Center of a Normal Distribution

☞ μ provides the center of the distribution– in fact, it is the mean.

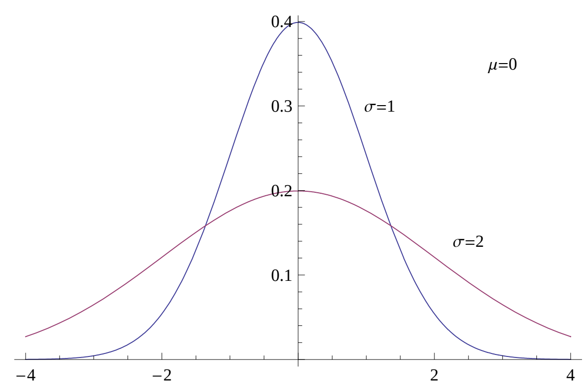


☞ The binomial and Poisson distributions approximate a bell curve – which is the graph of the density of a normal distribution. There is a deeper reason for this – more to follow.

Spread of a Normal Distribution

☞ σ^2 provides the spread of the distribution – in fact, it is the variance, where σ is the standard deviation.

The graphs show varying standard deviation σ.



☞ Note that the peak value, at $\mu = 0$, is $\frac{1}{\sqrt{2\pi}\sigma}$, where $\frac{1}{\sqrt{2\pi}} \approx 0.399$.

Standard Distribution

☞ There is only one normal distribution you need to know.

Defintion. The normal random variable with parameters $\mu = 0$ and $\sigma^2 = 1$ is called the standard normal random variable, which I will write as Z (as does Ross).

$$f_Z(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \qquad \text{for every } t \in \mathbb{R}.$$

The cumulative distribution for Z is written by Φ:

$$F_Z(a) = \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2}\, dt$$

Proof of equivalence

☞ If X is a normal random variable with parameters μ and σ^2, then

$$X = \sigma Z + \mu \qquad Z = \frac{X - \mu}{\sigma}.$$

Reason. Define g by

$$g(t) = \sigma t + \mu \qquad g^{-1}(t) = \frac{t - \mu}{\sigma}$$

Let $X = g(Z)$. By the previous theorem

$$\begin{aligned} f_X(t) &= f_Z\left(\frac{t-\mu}{\sigma}\right)\frac{1}{\sigma} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{t-\mu}{\sigma})^2/2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/2\sigma^2} \end{aligned}$$

Standardization

☞ To calculate the cumulative distribution for a normally distributed r.v. X with parameters μ and σ^2.
Reduce to the standard distribution:

$$\begin{aligned} F_X(a) &= \mathbf{P}\{X \le a\} \\ &= \mathbf{P}\{\sigma Z + \mu \le a\} \\ &= \mathbf{P}\{Z \le \frac{a-\mu}{\sigma}\} \\ &= \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

This process of changing a normal random variable to a standard one is known as standardization.

☞ The distribution Φ can be found in a table of values (see Ross, page 222).

Example

Example. Suppose X is normally distributed with parameter $\mu = 10$ and $\sigma^2 = 9$. What is $\mathbf{P}\{4 \le X \le 16\}$.

Solution. By standardizing X: $Z = (X - 10)/3$:

$$\begin{aligned} \mathbf{P}\{4 \le X \le 16\} &= \mathbf{P}\{\frac{4-10}{3} \le \frac{X-10}{3} \le \frac{16-10}{3}\} \\ &= \Phi(2) - \Phi(-2) \\ &= 2 \cdot \Phi(2) - 1 \approx 0.9544 \end{aligned}$$

☞ Typically, a table gives only values of $\Phi(a)$ for $a > 0$.

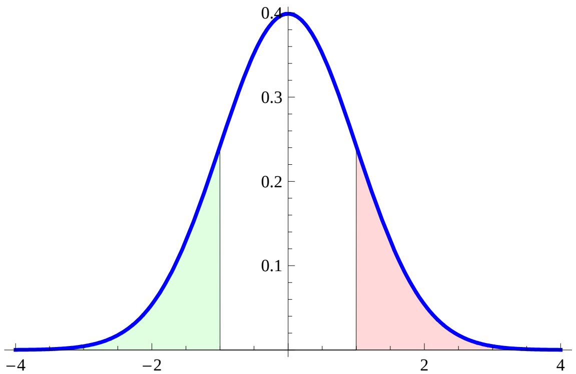
$$\Phi(-a) = 1 - \Phi(a).$$

When $a < 0$, so that $-a > 0$:
$\Phi(2) = 0.9772$ and $\Phi(-2) = 1 - \Phi(2) = 0.0228$.

Symmetry Normal Distribution

☞ The standard normal distribution is symmetric around $\mu = 0$.

$$\mathbf{P}\{Z < -1\} = 1 - \mathbf{P}\{Z < 1\} = \mathbf{P}\{1 < Z\}.$$



Symmetry of Normal Distribution

☞ The normal distribution Φ is symmetric about $\mu = 0$.

$$\Phi(-a) = 1 - \Phi(a) \qquad -\infty < a < \infty.$$

Reason. Key is to use change of variables (line 2)

$$\begin{aligned}
\Phi(-a) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-a} e^{-t^2/2}\,dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{a} -e^{-(-u)^2/2}\,du \qquad \text{let } u = -t \\
&= \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-u^2/2}\,du \\
&= \mathbf{P}\{a < Z\} = 1 - \mathbf{P}\{Z \le a\} \\
&= 1 - \Phi(a).
\end{aligned}$$

Three little facts

☞ This section is dedicated to proving three facts about the standard normal random variable Z.

1. $f_Z(t)$ is a probability density function.
2. $E[Z] = 0$.
3. $Var(Z) = 1$.

☞ It follows that for any normal random variable X with parameters μ and σ^2

1. $f_X(t)$ is a probability density function.
2. $E[X] = \mu$.
3. $Var(X) = \sigma^2$.

Standard density functions

❶ Let Z be the standard r.v. The function

$$f_Z(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

is a probability density function.

(a). It is clear that $f_Z(t) \ge 0$ for all real numbers t.

(b). We must show

$$1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2}\,dt$$

The problem is that the function $\int e^{-t^2/2}$ cannot be evaluated in terms of elementary functions, like you can with other familar functions from Calculus.

Standard density functions

☞ Instead we compute I, where

$$I = \int_{-\infty}^{\infty} e^{-t^2/2},$$

by computing I^2 and converting to double integral form

$$\begin{aligned} I^2 &= \Big(\int_{-\infty}^{\infty} e^{-t^2/2}\,dt\Big)\Big(\int_{-\infty}^{\infty} e^{-u^2/2}\,du\Big) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(t^2+u^2)/2}\,dt\,du \end{aligned}$$

☞ Convert to polar coordinates (the sum of squares in the exponents strongly suggest this), using the conversion

$$dt\,du = r\,dr\,d\theta \qquad \text{where } 0 \le \theta < 2\pi, 0 \le r < \infty$$
$$r = \sqrt{t^2 + u^2}$$

Standard density functions

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(t^2+u^2)/2}\,dt\,du \\ &= \int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2/2} r\,dr\,d\theta \\ &= 2\pi\int_{0}^{\infty} e^{-r^2/2} r\,dr \qquad s = -r^2/2,\ ds = -r\,dr \\ &= 2\pi\Big(-e^{-r^2/2}\Big)\Big|_0^{\infty} = 2\pi \end{aligned}$$

So, $I^2 = 2\pi$, or equivalently $I = \sqrt{2\pi}$.

☞ Therefore,

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-t^2/2}\,dt = \frac{1}{\sqrt{2\pi}} I = 1$$

Arbitrary density functions

❶ For any normal random variable X with parameters μ and σ^2, $f_X(t)$ is a density function.

☞ We have $X = \sigma Z + \mu$; standardize X by $Z = \frac{X-\mu}{\sigma}$.

$$f_X(t) = \frac{1}{\sigma} f_Z\Big(\frac{X-\mu}{\sigma}\Big)$$

Since $\sigma > 0$,

$$(a) \qquad 0 \le f_X(t) \qquad -\infty < t < \infty$$

$$\begin{aligned} (b) \quad \int_{-\infty}^{\infty} f_X(t)\,dt &= \int_{-\infty}^{\infty} \frac{1}{\sigma} f_Z\Big(\frac{t-\mu}{\sigma}\Big)\,dt \\ &= \int_{-\infty}^{\infty} f_Z(u)\,du \qquad u = \frac{t-\mu}{\sigma} \\ &= 1 \end{aligned}$$

Expectation for Standard Normal Distribution

❷ $E[Z] = 0$.

$$\begin{aligned} E[Z] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} te^{-t^2/2}\,dt \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -e^{u}\,du \qquad u = \frac{-t^2}{2} \\ &= -\frac{1}{\sqrt{2\pi}} e^{-t^2/2}\Big|_{-\infty}^{\infty} \\ &= \lim_{M\to\infty}\Big[\frac{e^{-(-M)^2/2}}{\sqrt{2\pi}} - \frac{e^{-M^2/2}}{\sqrt{2\pi}}\Big] \\ &= 0. \end{aligned}$$

Variance for Standard Normal Distribution

❷ $Var(Z) = 1$.

$$Var(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/2}\, dt$$

Integration by parts with ($u = t$ and $dv = te^{-t^2/2}$), so $v = -e^{-t^2/2}$,

$$\begin{aligned} Var(Z) &= \frac{1}{\sqrt{2\pi}} \left[-te^{-t^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-t^2/2}\, dt \right] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2}\, dt = 1 \end{aligned}$$

The second line uses the fact that

$$\lim_{M\to\infty} \left[-\frac{M}{e^{-M^2/2}} - \frac{M}{e^{-(-M)^2/2}} \right] = \lim_{M\to\infty} \frac{-2M}{e^{M^2/2}} = 0.$$

Arbitrary expectation and variance

☞ For any normal random variable X with parameters μ and σ^2,
❷ $E[X] = \mu$ and ❸ $Var(X) = \sigma^2$.

☞ We have $X = \sigma Z + \mu$.
Since $E[Z] = 0$ and $Var(Z) = 1$,

$$\begin{aligned} E[X] &= E[\sigma Z + \mu] = \sigma E[Z] + \mu = \mu \\ Var(X) &= Var(\sigma Z + \mu) = \sigma^2 Var(Z) = \sigma^2. \end{aligned}$$

Example: Grading on a curve

Example

Final exams at Podunk U. are constructed so that the distribution of scores is approximately normally distributed, with parameters μ (the average score) and σ (the standard deviation from the average). Letter grades are then assigned according to the following chart:

Test Score	Grade
$\mu + \sigma < x$	A
$\mu < x < \mu + \sigma$	B
$\mu - \sigma < x < \mu$	C
$\mu - 2\sigma < x < \mu - \sigma$	D
$x < \mu - 2\sigma$	F

☞ This system of assigning letter grades is called “grading on the curve".

Example: Grading on a curve

☞ Let X be a normally distributed r.v. with parameters μ and σ.
By standardization: $Z = \frac{X-\mu}{\sigma}$.

$$\begin{aligned} \mathbf{P}\{\mu + \sigma < X\} &= \mathbf{P}\{1 < \frac{X-\mu}{\sigma}\} = 1 - \Phi(1) \approx 0.1587 \\ \mathbf{P}\{\mu < X < \mu + \sigma\} &= \mathbf{P}\{0 < \frac{X-\mu}{\sigma} < 1\} = \Phi(1) - \Phi(0) \approx 0.3413 \\ \mathbf{P}\{\mu - \sigma < X < \mu\} &= \mathbf{P}\{-1 < \frac{X-\mu}{\sigma} < 0\} \\ &= \Phi(0) - \Phi(-1) = \Phi(0) + \Phi(1) - 1 \approx 0.3413 \\ \mathbf{P}\{\mu - 2\sigma < X < \mu - \sigma\} &= \mathbf{P}\{-2 < \frac{X-\mu}{\sigma} < -1\} \\ &= \Phi(-2) - \Phi(-1) = \Phi(1) - \Phi(2) \approx 0.1359 \\ \mathbf{P}\{X < \mu - 2\sigma\} &= \mathbf{P}\{\frac{X-\mu}{\sigma} < -2\} = 1 - \Phi(2) = 0.0228 \end{aligned}$$

The probabilities can be computed from a table for the standard normal curve.