**Stat 100a: Introduction to Probability.**

Outline for the day:

1. Continuous random variables and densities.
2. Exponential random variables.
3. Uniform, random variables.
4. Normal random variables.
5. Moment generating functions of different distributions.
6. Survivor functions.
7. Midterm one.

Read through chapter 6.4!

♣ ♥ ◆ ♠

## 1. Continuous random variables and their densities, p103-107.

Density (or pdf = Probability Density Function) f(y):

$$\int_B f(y)\,dy = P(X \text{ in } B).$$

Expected value, $\mu = E(X) = \int y\, f(y)\, dy$. ($= \sum y\, P(y)$ for discrete X.)

Variance, $\sigma^2 = V(X) = E(X^2) - \mu^2$.

$SD(X) = \sqrt{V(X)}$.

For examples of pdfs, see p104, 106, and 107.

**2. Exponential distribution, ch 6.4.**

Useful for modeling waiting times til something happens (like the geometric).

pdf of an exponential random variable is $f(y) = \lambda\ exp(-\lambda\ y)$, for $y \geq 0$,
and $f(y) = 0$ otherwise.
The cdf is $F(y) = 1 - exp(-\lambda\ y)$, for $y \geq 0$.
If $X$ is exponential with parameter $\lambda$, then $E(X) = SD(X) = 1/\lambda$

If the total numbers of events in any disjoint time spans are independent, then these totals are Poisson random variables. If in addition the events are occurring at a constant rate $\lambda$, then the times between events, or *interevent times*, are exponential random variables with mean $1/\lambda$.

**Example.** Suppose you play 20 hands an hour, with each hand lasting exactly 3 minutes, and let $X$ be the time in hours until the end of the first hand in which you are dealt pocket aces. Use the exponential distribution to approximate $P(X \leq 2)$ and compare with the exact solution using the geometric distribution.

**Answer.** Each hand takes 1/20 hours, and the probability of being dealt pocket aces on a particular hand is 1/221, so the rate $\lambda$ = 1 in 221 hands = 1/(221/20) hours ~ 0.0905 per hour.
Using the exponential model, $P(X \le 2 \text{ hours}) = 1 - exp(-2\lambda) \sim 16.556\%$.
This is an approximation, however, since by assumption $X$ is not continuous but must be an integer multiple of 3 minutes.
Let $Y$ = the number of hands you play until you are dealt pocket aces. Using the geometric distribution, $P(X \le 2 \text{ hours}) = P(Y \le 40 \text{ hands})$
$= 1 - (220/221)^{40} \sim 16.590\%$.

The survivor function for an exponential random variable is particularly simple: $P(X > c) = \int_c^\infty f(y)dy = \int_c^\infty \lambda\, exp(-\lambda\, y)dy = -exp(-\lambda\, y)]_c^\infty = exp(-\lambda\, c)$.

Like geometric random variables, exponential random variables have the *memorylessness* property: if $X$ is exponential, then for any non-negative values $a$ and $b$, $P(X > a+b \mid X > a) = P(X > b)$. *(See p115).*
Thus, with an exponential (or geometric) random variable, if after a certain time you still have not observed the event you are waiting for, then the distribution of the *future,* additional waiting time until you observe the event is the same as the distribution of the *unconditional* time to observe the event to begin with.

## 3. Uniform Random Variables and R.

Continuous random variables are often characterized by their

*probability density functions* (pdf, or *density*): a function f(x)

such that P{X is in B} = $\int_B f(x)\,dx$.

Uniform: f(x) = c, for x in (a, b).

= 0, for all other x.

[Note: c must = 1/(b-a), so that $\int_a^b f(x)\,dx$ = P{X is in (a,b)} = 1.]

Uniform (0,1). See p107-109.

f(y) = 1, for y in (0,1). $\mu = 0.5$. $\sigma \sim 0.29$.

P(X is between 0.4 and 0.6) = $\int_{.4}^{.6} f(y)\,dy = \int_{.4}^{.6} 1\,dy = 0.2$.

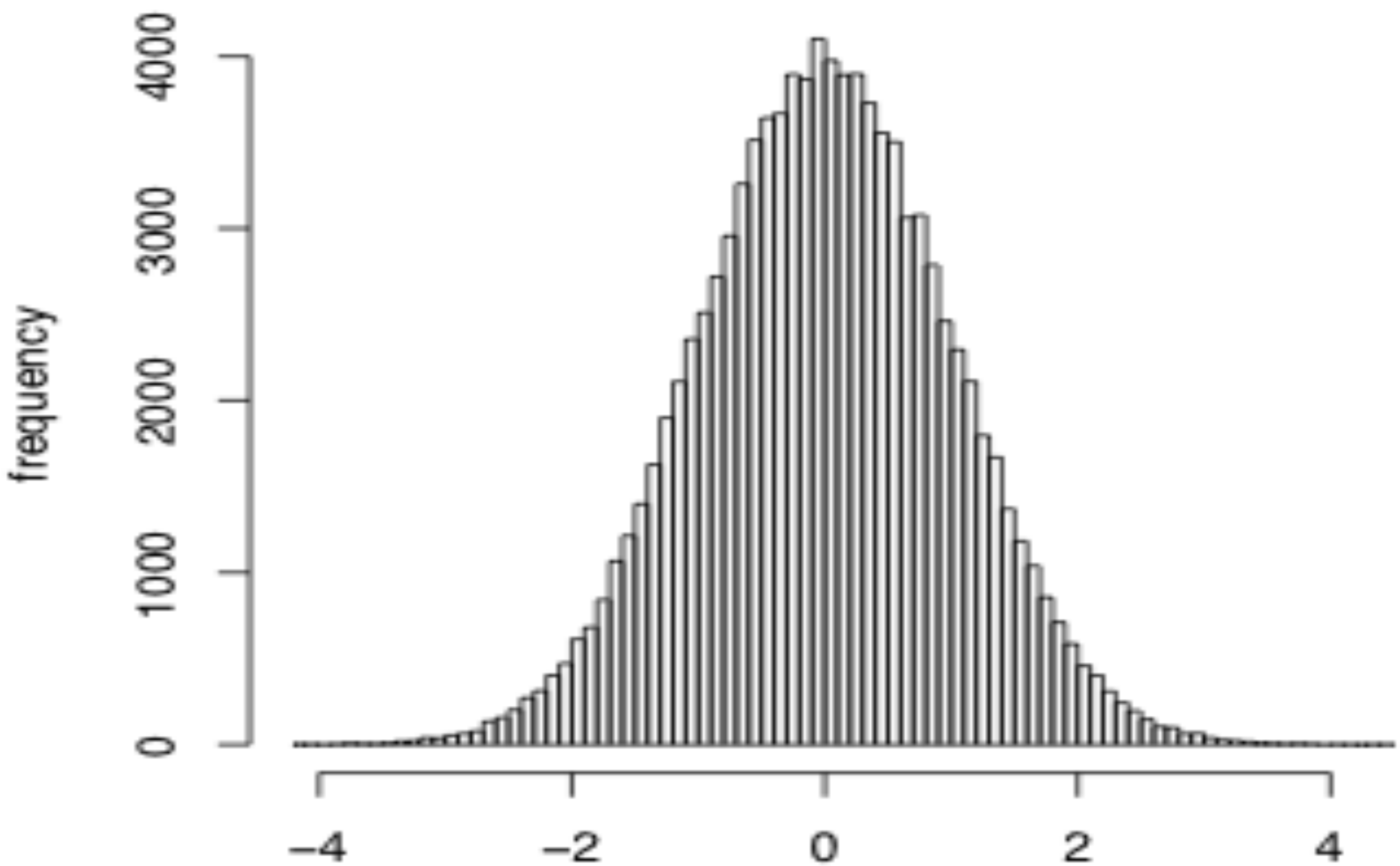
In R, runif(1,min=a,max=b) produces a pseudo-random uniform.

**4. Normal random variables,** pp 115-117.

mean = $\mu$, SD = $\sigma$, $f(y) = 1/\sqrt{(2\pi\sigma^2)}\ e^{-(y-\mu)^2/2\sigma^2}$.
Symmetric around $\mu$,
50% of the values are within 0.674 SDs of $\mu$,
68.27% of the values are within 1 SD of $\mu$, and
95% are within 1.96 SDs of $\mu$.

* Standard Normal. Normal with $\mu = 0$, $\sigma = 1$. See pp 117-118.

Standard normal density:
68.27% between -1.0 and 1.0
95% between -1.96 and 1.96



## 5. Moment generating functions of some random variables.

Bernoulli(p). $ø_X(t) = pe^t + q$.

Binomial(n,p). $ø_X(t) = (pe^t + q)^n$. p94.

Geometric(p). $ø_X(t) = pe^t/(1 - qe^t)$.

Neg. binomial (r,p). $ø_X(t) = [pe^t/(1 – qe^t)]^r$. p97.

Poisson($\lambda$). $ø_X(t) = e^{\{\lambda e^t - \lambda\}}$. p100.

Uniform (a,b). $ø_X(t) = (e^{tb} – e^{ta})/[t(b-a)]$. p108.

Exponential ($\lambda$). $ø_X(t) = \lambda/(\lambda-t)$. p123.

Normal. $ø_X(t) = e^{\{t\mu + t^2\sigma^2/2\}}$.

**6. Survivor functions. p96 and 115.**

Recall the cdf $F(b) = P(X \leq b)$.
The survivor function is $S(b) = P(X > b) = 1 - F(b)$.
Some random variables have really simple survivor functions and it can be convenient to work with them.
If X is geometric, then $S(b) = P(X > b) = q^b$, for b = 0,1,2,....
For instance, let b=2. X > 2 means the 1st two were misses,
i.e. $P(X>2) = q^2$.
For exponential X, $F(b) = 1 - exp(-\lambda b)$, so $S(b) = exp(-\lambda b)$.

An interesting fact is that, if X takes on only values in {0,1,2,3,...},
then $E(X) = S(0) + S(1) + S(2) + ....$
Proof. See p96.

$$
\begin{aligned}
S(0) &= P(X=1) + P(X=2) + P(X=3) + P(X=4) + .... \\
S(1) &= \phantom{P(X=1) + {}} P(X=2) + P(X=3) + P(X=4) + .... \\
S(2) &= \phantom{P(X=1) + P(X=2) + {}} P(X=3) + P(X=4) + .... \\
S(3) &= \phantom{P(X=1) + P(X=2) + P(X=3) + {}} P(X=4) + ....
\end{aligned}
$$

Add these up and you get
$0\, P(X=0) + 1P(X=1) + 2P(X=2) + 3P(X=3) + 4P(X=4) + ...$
$= \sum kP(X=k) = E(X)$.