Lecture #10
6/7/11

Admin

- usual stuff about HW
- midterm review session

Plan

- Cont r.v's — ch 12
- exp/var analogues
- bell curve / normal curve
- CLT
- "empirical rule" (end of ch 4)
- Standardization
- examples
- Q-Q plot

Yesterday, we did cont. r.v.'s ...

$f_X(x)$ ← PDF

$\text{support}[X] = [0, 10]$ any # in there

$$P\left(X \in [2,7]\right) = \int_2^7 \frac{1}{10}\,dx = \frac{1}{10}[x]_2^7 = \frac{1}{10}(7-2) = \frac{5}{10} = \frac{1}{2}$$

Discrete r.v.:

$$\mu = E[X] = \sum_{i=1}^{K} x_i P(X=x_i)$$

$$\sigma^2 = \mathrm{Var}[X] = \sum_{i=1}^{K} (x_i-\mu)^2 P(X=x_i)$$

Cont. r.v.

$$\mu = E[X] = \int_{-\infty}^{\infty} x\, f_X(x)\,dx$$

$$\sigma^2 = \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx$$

Many different types of cont. r.v.'s. We're only going to discuss the normal r.v.

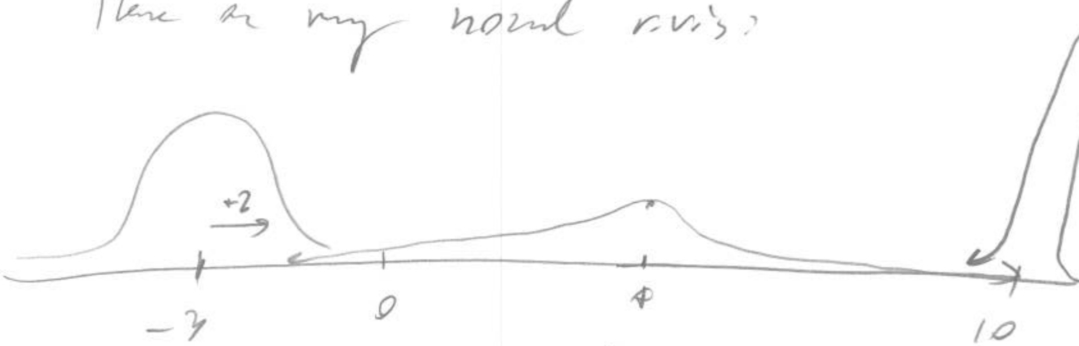
mean, variance

$$X \sim N(\mu, \sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad Z \sim N(0,1) \triangleq \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

general normal r.v. std normal r.v.

exercise: $E[Z] = \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 0$

There are many normal r.v.'s:



$X \sim N(-3, 1)$ $Q \sim N(4, 2^2)$ $Y \sim N(10, (\frac{1}{3})^2)$

Shifts and Scales...

Who cares...

be patient...

But... if X is normal, then

Z is a special letter

"standardization" $$\frac{X-\mu}{\sigma} \overset{d}{=} Z \sim N(0,1)$$

Why $E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}E[X-\mu] = \frac{1}{\sigma}(E[X]-\mu) = 0$ ✓

$Var[Z] = Var\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma^2}Var[X-\mu] = \frac{1}{\sigma^2}Var[X] = 1$ ✓

Wool over your eyes: multiplying or adding a const to a normal still keeps it normal: not proved...

Let's find some basic probs. ~~the~~

- What is the prob that $Z \geq 0$?

$$P(Z \geq 0) = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz$$

This is easy... just get the integral and do it...

turns out... that integral is impossible. It cannot be done! The only way to do it is numerical approximations i.e. Riemann sums (or something similar)

(Pass out tables now)

These tables have done the work for you. If you don't recognize them, they're the back page of S&F.

You should carry around these tables everywhere...
like your Penn ID...

4

Using the table... what is $P(Z \geq 0)$? Does this make sense?

How about $P(Z \geq 2 \text{ OR } Z \leq -2)$

$= P(|Z| \geq 2)$

$= ?$

How about $P(Z \leq 1 \text{ AND } Z \geq -1)$

$= P(|Z| \leq 1)$

$\approx .68$

$P(Z \leq 2 \text{ AND } Z \geq -2)$

$= P(|Z| \leq 2)$

$\approx .95$

$P(Z \leq 3 \text{ AND } Z \geq -3)$

$= P(|Z| \leq 3)$

$\approx .997$

Lots of the problems using the table

ch4 p64/65 Empirical Rule... if the density/histogram is about bell-shaped then in between ± 1 SD $\Rightarrow$ 68%, ± 2 SD $\Rightarrow$ 95% and ± 3 SD $\Rightarrow$ 99.7%

Okay... who cares...? Demo time

Prove in discrete / cont / weird case...

What did we find?

$X_1, \ldots, X_n \overset{iid}{\sim}$ something discrete or cont $\Rightarrow S_n = X_1 + \ldots + X_n \sim N(\cdot, \cdot)$

AND " " " $\Rightarrow \bar{X}_n \sim N(\cdot, \cdot)$ as well

Central Limit Theorem

Sums of iid r.v.'s ~~are~~ are approx normal if n is large
Eventually it will get there...

QUOTE

How large is large? For purposes of this class $n \geq 30$. Ignore bottom p328 sample size conditions

Sir Francis Galton (*Natural Inheritance*, 1889) described the Central Limit Theorem as:

> " I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshaled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

Let's derive the two major formulas...

$S_n = X_1 + \dots + X_n$, $E[S_n] = n\mu$, $Var[S_n] = n\sigma^2$

if n large $\Rightarrow S_n \sim N\left(n\mu, \overset{SD}{(\sqrt{n}\sigma)}^2\right)$

$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$, $E[\bar{X}_n] = \mu$, $Var(\bar{X}_n) = \frac{\sigma^2}{n}$

if n large $\Rightarrow \bar{X}_n \sim N\left(\mu, \overset{SD}{\left(\frac{\sigma}{\sqrt{n}}\right)}^2\right)$

You spend \$30 ± 10 at the grocery store once per week.

What's the prob. you spend more than \$1800 at the grocery store next year?

$S_n \sim N(n\mu, (\sqrt{n}\sigma)^2) = N(1560, 72.11^2)$

$P(S_n \geq 2000) = P\left(\frac{S_n - 1560}{72.11} \geq \frac{1800 - 1560}{72.11}\right) = P(Z > 3.32) = \boxed{.0004}$

Next Q: 10% of the time, I would spend more than ...

Normal Q-Q plots if time...