Review (Ch. 1 – 3.1 inclusive)

Ch1 Data : cases , variables.

types of variables : categ. / continuous.

Sample vs. population :

describe with distribution

describe with histograms, etc.

sample

histograms ; freq., rel. freq., density.

Area ~ long-run proportion.

pop.

distribution

mass / density function.

$p(x)$ $f(x)$.

$\|$ $\shortparallel$

Categ. Cont.

Example of $p(x)$:

Binomial : $p(x) = \binom{n}{x} \pi^{x} (1-\pi)^{n-x}$

$$x = 0, 1, \ldots, n.$$

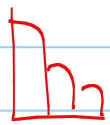
In the limit $n \to \infty$, $\pi \to 0$, $n\pi$ = const. = λ

Binomial $\longrightarrow$ poisson

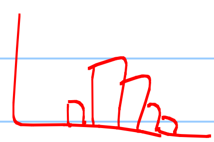
$$p(x) = \binom{n}{x} \pi^{x} (1-\pi)^{n-x} \longrightarrow p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

x = # of H's in n tosses

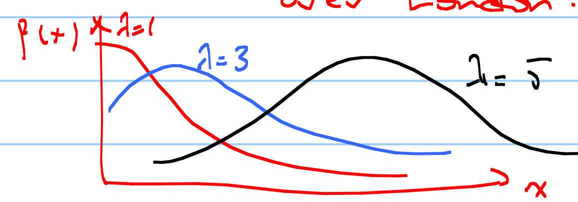
x = # of bombs per block over London.



smaller n



larger n



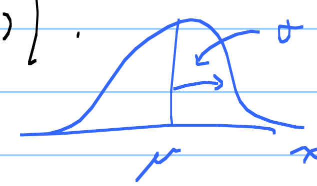
In the larger n limit (mid-range π)

binomial $\longrightarrow$ normal

$$n, \pi \longrightarrow \mu = n\pi, \quad \sigma^2 = n\pi(1-\pi).$$

Normal distr. is for continuous x. ie.

it's an example of $f(x)$ { not $p(x)$ }.

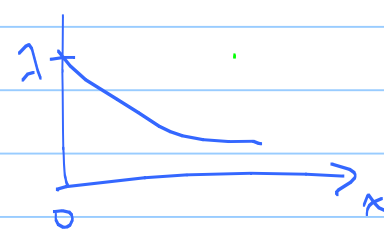


$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \quad -\infty < x < \infty$$

With $\mu = 0$, $\sigma = 1$, it's called std. Normal.

Another example of $f(x)$ is Exponential:

$$f(x) = \lambda e^{-\lambda x} \qquad x \geq 0.$$



Note all of these satisfy

$$p(x) \geq 0 \qquad \sum_x p(x) = 1$$

$$f(x) \geq 0 \qquad \int f(x)\, dx = 1.$$

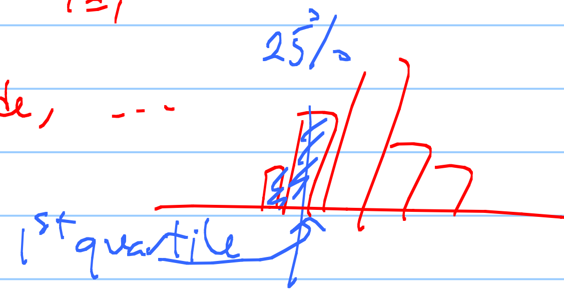
And so, area $\sim$ long-run proportion.

(Ch2) Summary Measures of data (and of dists)

↑ general and special.

location:

- sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- " median, mode, ...
- " quartile



spread:

- " std. dev. $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
- " range = max − min.
- " IQR = 3rd qrt. − 1st qrt.

5-number summary : box plot.

Also, recall computational formula for s^2

$$s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 = \frac{n}{n-1}\left(\overline{x^2} - \bar{x}^2\right)$$

Now, summary measures for distr.

mean : $\mu_x = E[x] = \sum x\, p(x)$, $\int x f(x)\, dx$.

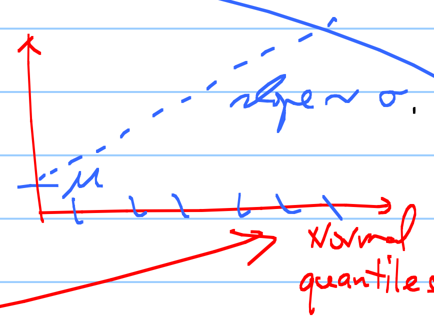
Variance : $\sigma_x^2 = V[x] = \sum (x-\mu_x)^2 p(x)$, ...

And more : median, mode, quartile, box plot, ...

computational formula for variance : Analogous to the sample version above.

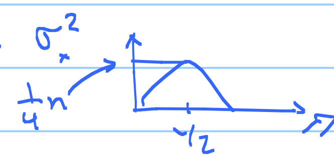
Normal qq-plots. NOT Covered on Test 1

Sample quantiles



$E[x]$, $V[x]$ for special distr. (above)

	$\mu_x = E[x]$	$\sigma_x^2 = V[x]$
Binomial	$n\pi$	$n\pi(1-\pi)$
Normal	μ	σ^2
poisson	λ	λ



Variance of binomial is constrained to be less than $n/4$.

Finally, by now you are ready to appreciate the following Table:

Analogy between probability and statistics.

probability	statistics
one coin (H/T)	one individual (B/G)
one toss	one selection.
n tosses	1 sample of size n.
π = prob. of H	π = prop. of Boys in pop
X = r.v. = # of H's in $\underline{n}$ Tosses	X = # of boys in a sample of size $\underline{n}$
$P(X=3, n=4) = \binom{4}{3} \pi^3 (1-\pi)^{4-3}$	prob. of 3 boys in a sample of size 4.
$E[x] = \underset{\text{sample size}}{n}\ \underset{\text{pop. param.}}{\pi}$	Avg. (over ∞ samples of size n) of The # of boys in samples of size n.
$V[x] = n\pi(1-\pi)$	Variance (Ibid) Ibid.