(c) (n fixed, $T \uparrow$) $\alpha^{\uparrow 1}$ $\leftrightarrow$ with a small sample from a large population,

SRS $\doteq$ IID

Poisson ($\lambda > 0$) $X \sim \text{Poisson}(\lambda)$

$\leftrightarrow$ X has PMF $f_X(x) = \dfrac{\lambda^x e^{-\lambda}}{x!} \underbrace{I_{\{0,1,\dots\}}(x)}_{\text{support of } X}$

$\boxed{E(X) = \lambda}$

$\boxed{V(X) = \lambda}$

$\Big\}$ thus for the Poisson dist.

$$\frac{V(X)}{E(X)} = 1$$

$\boxed{\text{Def.}}$ If $E(X)$ and $V(X)$ both exist and $E(X) \neq 0$, $\dfrac{V(X)}{E(X)}$ is called the variance-to-mean ratio (VTMR)

$\psi_X(t) = e^{\lambda(e^t - 1)}$

$-\infty < t < \infty$

The Poisson can be unrealistic as a consequence of its VTMR of 1, $\to$ because many rvs that represent counts of occurrences of events in time intervals of fixed length have VTMR > 1.

The Poisson & Binomial distributions both count the numbers of "successes" in a process unfolding in time, so it should not be surprising to find out that these 2 dist. are related:

---

when $\begin{pmatrix} n \text{ is large} \\ p \text{ is close to } 0 \end{pmatrix}$, $\text{Binomial}(n, p) \doteq \text{Poisson}(n \cdot p)$

Theorem: n positive integer, $0 < p < 1$, $X \sim \text{Binomial}(n, p)$

$\lambda > 0$, $Y \sim \text{Poisson}(\lambda)$ / Choose any sequence

$\{p_n\}_{n=1}^{\infty}$ of values between 0 and 1 with

$\lim_{n \to \infty} n \cdot p_n = \lambda$ Then $f_X(x \mid n, p_n) \underset{n\to\infty}{\longrightarrow} f_Y(y \mid \lambda)$

Poisson process, revisited

Def

A poisson process with rate λ per unit time (or space, or volume, or...) is a stochastic process with two properties:

(a) # arrivals in every interval of time of length $t \sim \text{Poisson}(\lambda t)$

(b) #s of arrivals in all disjoint (non-overlapping) time intervals are independent

~~Case Study~~

~~Parasitic~~ protozoa in drinking water

There's a kind of parasitic

organism called <u>cryptosporidium</u> that's capable of getting into the public drinking water supplies; at one stage in their life cycle they're called <u>oocysts</u>. They can make people sick at a concentration of only 1 oocyst per 5 liters $\doteq$ 1.3 gallons of water

One problem is that it can be hard to detect these oocysts with water filtration. Suppose that, in the water supply of your city, oocysts occur according to a Poisson process with rate λ oocysts per liter, & that the filtering system your water utility company uses can capture all the oocysts in a water sample but only has probability p of detecting each oocyst that's actually there. (& counting events are independent)

Set Y = # actual oocysts in t liters of water, and $X_i = \begin{cases} 1 & \text{if oocyst } i \text{ gets counted} \\ 0 & \text{else} \end{cases}$

X = # counted oocysts | Then $(X \mid Y = y) = \sum_{i=1}^{y} X_i$

Under these assumptions, $(X \mid Y = y) \sim \text{Binomial}(y, p)$

Q: What's the dist. of X? | A | By the law of total probability

$$f_X(x) = P(X = x) = \sum_{y=0}^{\infty} P(Y = y) P(X = x \mid Y = y)$$

for all $x = 0, 1, \ldots$

in which $P(Y = y) = \dfrac{(\lambda t)^y e^{-\lambda t}}{y!}$ for $y = 0, 1, \ldots$

and $P(X = x \mid Y = y) = \binom{y}{x} p^x (1-p)^{y-x}$

Notice that if $X = x$, $Y \geq x$ because the actual number of oocysts (Y) has to be at least as large as the number of oocysts detected (X). After a careful calculation you get,

$$f_X(x) = \sum_{y=x}^{\infty} \binom{y}{x} p^x (1-p)^{y-x} \frac{(\lambda t)^y e^{-\lambda t}}{y!}$$

$$= \frac{e^{-p\lambda t} (p\lambda t)^x}{x!}$$

i.e., $X \sim \text{Poisson}(p\lambda t)$: losing a proportion $(1-p)$ of the oocysts to faulty counting just lowers the rate of the Poisson process from λ / liter to $\lambda \cdot p$ / liter (makes excellent sense).

In practice oocysts are hard to detect: p is small (not far from 0). **Q:** How much water (t liters) do you need to filter to achieve $P(\text{at least 1 oocyte detected}) \geq 1-\alpha$ for small α? **A:** Not hard to work out

$$P(\text{at least 1 detected}) = 1 - P(\text{none detected})$$

$$= 1 - P(X = 0) = 1 - e^{-p\lambda t} \geq 1-\alpha$$

$$\leftrightarrow \alpha \geq e^{-p\lambda t} \leftrightarrow \ln\alpha \geq -p\lambda t \leftrightarrow$$

$$\boxed{t \geq -\frac{\ln\alpha}{p\lambda}}$$

Example $\alpha = .01$, $p = 0.1$, $\lambda = 0.2$/liter (1 per 5 liters)

minimum sickness level

to achieve prob 99%, t has to be at least $\underline{230.3}$ liters.

Negative Binomial Distribution

You're watching a potentially endless sequence of Bernoulli trials with constant success probability p.

Let X = # failures before rth success

r integer ≥ 1

You can show that X follows what's called the Negative Binomial dist:

its PMF is $f_X(x \mid r, p) = \binom{r+x-1}{x} p^r (1-p)^x \, I_{\{0,1,2,\dots\}}(x)$.

with parameters (r, p) $(0 < p < 1)$

The name comes from the fact that, when you watch a sequence of Bernoulli trials with constant unknown success probability p unfold, there are two different ways to estimate p: decide ahead of time to sample n (known constant) success/failure trials, and record the (random) # S of successes you see (from which a reasonable estimate would be $\hat{p}_B = \frac{S}{n}$ ← Binomial);

---

or decide ahead of time that you're going to sample until you've seen s (known constant) successes & record the (random) # of trials N needed to accumulate that many successes (from which a reasonable estimate would be $\hat{p}_{NB} = \frac{s}{N}$ ← Negative Binomial).

Special case of Negative Binomial

Set $r=1$ and record the number $\mathbb{X}$ of failures until the first success: $\mathbb{X}$ is said to follow the

Geometric (p) distribution, with

PMF $f_{\mathbb{X}}(x \mid p) = p(1-p)^x \, \underbrace{I_{\{0,1,\ldots\}}(x)}_{\text{support of } \mathbb{X}}$.

(parameter p)

Consequence $\mathbb{X}_1, \ldots, \mathbb{X}_n$ IID Geometric(p)

$\rightarrow \sum_{i=1}^{n} \mathbb{X}_i \sim$ Negative Binomial (n, p)

This is a direct analogue to the Bernoulli/Binomial story: $\mathbb{X}_1, \ldots, \mathbb{X}_n$ IID Bernoulli (p) $\rightarrow \sum_{i=1}^{n} \mathbb{X}_i \sim$ Binomial (n, p).

$\boxed{X \sim \text{Negative Binomial } (r, p)}$

$$\psi_X(t) = \left[\frac{p}{1-(1-p)e^t}\right]^r \text{ for } t < \log\left(\frac{1}{1-p}\right)$$

from which $E(X) = \dfrac{r(1-p)}{p}$, $V(X) = \dfrac{r(1-p)}{p^2}$

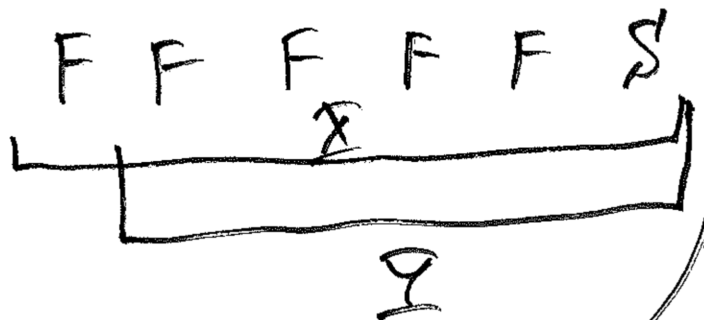
Consequence | $X \sim \text{Geometric}(p) \to$

$\left\{\begin{matrix} k \\ t \end{matrix}\right\}$ both non-negative integers

$$P(X = k+t \mid X \geq k) = P(X = t)$$

this is called the <u>memoryless property</u> of the Geometric distribution, and it turns out that this is the <u>only</u>

discrete distribution with this property.



X = # failures until first success = 5 (here)

Y = # failures, starting at trial ($k+1 = 2$ here), until next success (= 4 here). Then Y has the same dist. as X and is independent of what happened on the first k trials, i.e., "the process has no memory".

Case 2: Important Continuous Distributions

Normal (Gaussian) Distribution

$X \sim \text{Normal}(\mu, \sigma^2)$ mean μ, variance $0 < \sigma^2 < \infty$

PDF

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

The Normal dist. is the single most important dist. in all of probability & statistics, mainly for 2 reasons: (1) many observable random processes have dist. shapes that are close to the "bell curve" (Normal PDF), and (2) the Central Limit Theorem (CLT), which we'll examine soon.

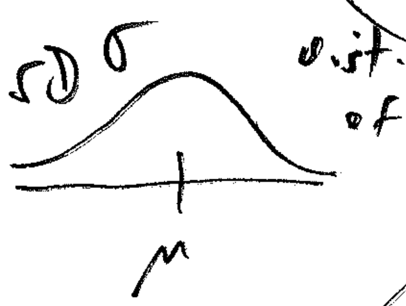
Properties of the Normal Dist.

$N(\mu, \sigma^2)$

$X \sim \text{Normal}(\mu, \sigma^2)$ $\quad E(X) = \mu$

$V(X) = \sigma^2, \quad SD(X) = \sigma$

$$\psi_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$



center of symmetry, mean, median, mode: all $= \mu$

Consequences | ① $X \sim \text{Normal}(\mu, \sigma^2)$,

$Y = aX + b$, $\left(\begin{matrix} a \neq 0 \\ b \end{matrix} \right)$ fixed constants $\longrightarrow$

$Y \sim \text{Normal}(a\mu + b, a^2\sigma^2)$.

In other words, Normality is preserved under linear transformation

Def.

The Normal dist. with mean $\mu = 0$ and SD $\sigma = 1$ is called the standard Normal dist.

The PDF of $X \sim \text{Normal}(0, 1)$ is

$$\phi_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$ and its

phi (lower-case)

CDF is $\Phi_X(x) = \int_{-\infty}^{x} \phi_X(t)\, dt$

uppercase Phi

23 Aug 17

It turns out that e^{-cx^2} $(c>0)$ has no anti-derivative in closed form, so $\Phi(x)$ cannot be summarized in a formula; instead it's approximated by numerical integration (see p 861 in DS).

Consequences, continued

② Because the Normal PDF (for all $x \in \mathbb{R}$) is symmetric, $\Phi(-x) = 1 - \Phi(x)$

and $\Phi^{-1}(p) = -\Phi^{-1}(1-p)$ (for all $0<p<1$)

③ $X \sim \text{Normal}(\mu, \sigma^2) \rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$,

so that $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

and $F_X^{-1}(p) = \mu + \sigma\, \Phi^{-1}(p)$

Empirical Rule

Part 1 start at the mean μ of a distribution and go 1 SD σ either way: you will find (about $\frac{2}{3}$) 68% of the probability in the interval $(\mu \pm 1\sigma)$

Part 2 Ditto 2 SDs either way: $(\mu \pm 2\sigma)$ captures most (about 95%) of the probability

Part 3 Ditto 3 SDs either way: $(\mu \pm 3\sigma)$ captures almost all 99.7% of the probability

This Rule is <u>exact</u> for all Normal dists & is a surprisingly good approximation for many other distributions. This permits an easy trick that's helpful in computing Normal probabilities.

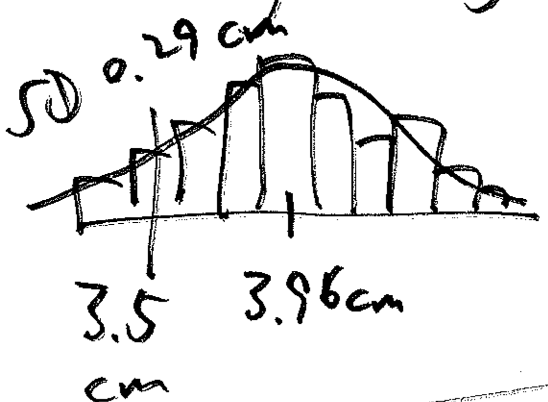
Example: You have a random sample of $n = 103$ immature monarch butterflies, and you measure their wing lengths:

y = wing length (cm)

$$\left[\begin{array}{l} y_1 = 4.1 \\ y_2 = 3.3 \\ \vdots \\ y_n = 4.7 \end{array}\right] \quad n = 103$$

mean $\bar{y} = 3.96$ cm

SD $s = 0.29$ cm



histogram of wing length

Q: About what % of the sampled butterflies had wing length ≤ 3.5 cm?

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \qquad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

sample mean sample SD

sorted y

3.2
3.3
⋮
3.5
3.5
3.5
3.5

3.6
⋮
4.7

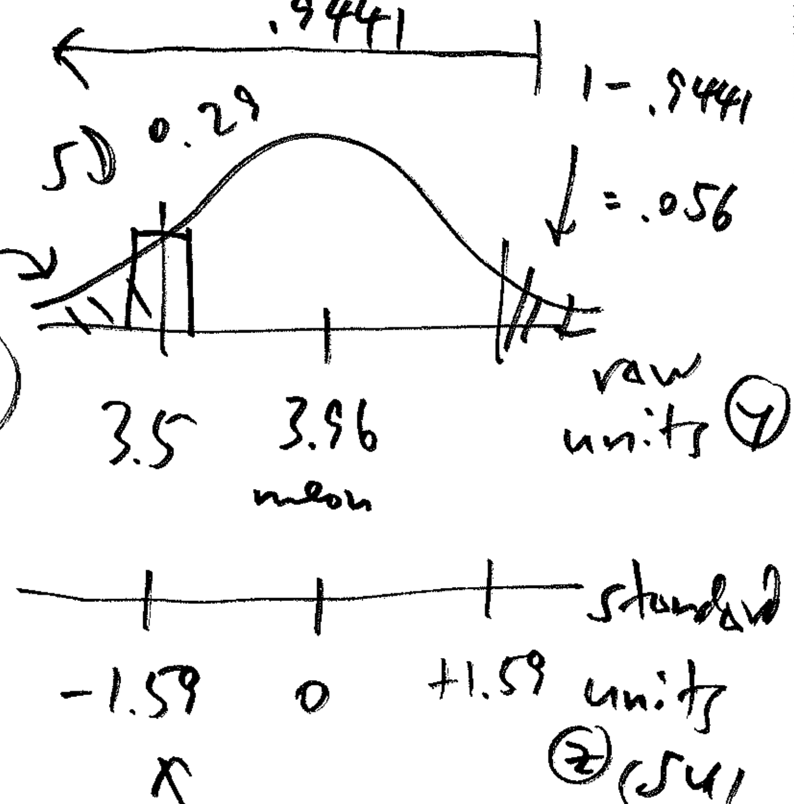
4

8

103

$\boxed{A_1}$ (exact) $\frac{8}{103} \doteq 7.8\%$ 263

$\boxed{A_2}$ (approximate)



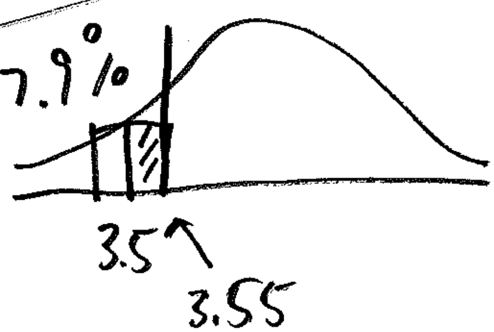
$$\frac{3.5 - 3.96}{0.29}$$

converting to SU for data:

$$z = \frac{y - \bar{y}}{s} = SU$$

for random variables

$$Z_Y = \frac{Y - \mu}{\sigma} = SU$$



keeping track of histogram bar edges: <u>continuity</u> <u>correction</u>

More consequences

(4) $X_1, \ldots, X_k$ independent,
$X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$

$$\to \sum_{i=1}^{k} X_i \sim \text{Normal}\left(\sum_{i=1}^{k}\mu_i, \sum_{i=1}^{k}\sigma_i^2\right)$$

nice additive property

this, is why Normal dist.s are indexed by variance rather than SD.

Notation: $\text{Normal}(\mu, \sigma^2) \triangleq N(\mu, \sigma^2)$

Example: Population of adult U.S. women: height follows $N(\mu = 65.0\text{in}, \sigma^2 = 3.2^2\text{in}^2)$ dist.
$(\sigma = 3.2\text{ in})$

Pop. of adult U.S. men: height follows $N(\mu = 69.5\text{ in}, \sigma^2 = 3.3^2\text{ in}^2)$ dist.

1 woman chosen at random, height $\underline{W}$;
1 man chosen at random (independently), height $\underline{M}$; P(woman taller than man)

$$= P(\underline{W} > \underline{M}) = ?$$

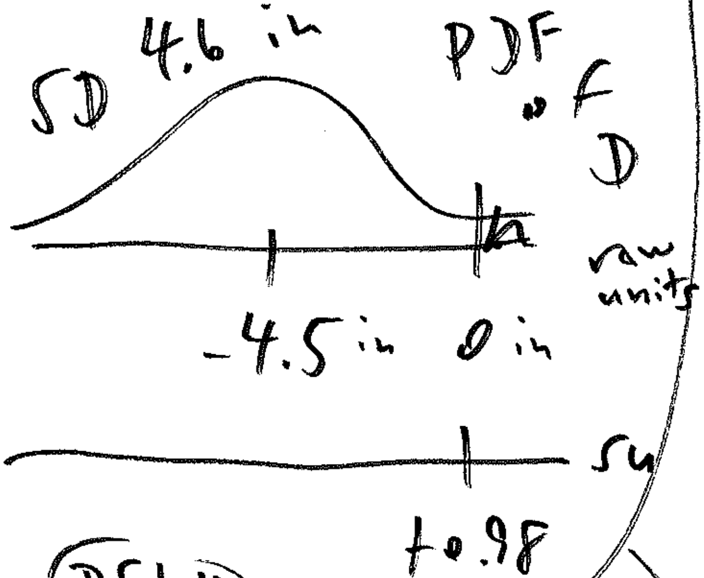
Define $D = \underline{W} - \underline{M}$

By consequence 4, $D \sim N(65 - 69.5 = -4.5$ in, $3.2^2 + 3.3^2 = 21.1$ in$^2)$

$\downarrow$

SD $\sqrt{21.1 \text{ in}^2} = 4.6$ in

$$P(\underline{W} > \underline{M}) = P(D > 0)$$



Convert to SU:

$$\frac{0 - (-4.5)}{4.6} = +0.98$$

DS table

$\leftarrow$ 0.8365 | .1635

So $P(\underline{W} > \underline{M}) \doteq 16\%$ (about 1 in 6)

Def rv $X_1, \ldots, X_n \to$ <u>sample mean</u> of $(X_1, \ldots, X_n)$ is $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

Consequences, continued

(5) $\left\{ X_i \overset{IID}{\sim} N(\mu, \sigma^2) \atop (i = 1, \ldots, n) \right\} \to \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ so $SD(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$

Because $E(\bar{X}_n) = \mu$, $\bar{X}_n$ is an <u>unbiased</u> <u>estimator</u> of μ

Def. In frequentist statistics, the standard deviation (SD) of an estimator $\hat{\theta}$ (rv) of a parameter θ is called the <u>standard error</u> $SE(\hat{\theta})$ of $\hat{\theta}$.