CSE 312

Foundations of Computing II

Section 7 : Continuity Correction

Deadlines/Announcements

- Midterm grades should be released soon (if not already!)
- Pset 6 released, due next Wednesday 2/21 at 11:59 pm

Review Exponential Distribution

X~Exp(λ) tells how much time till a certain event happens
(λ *is the rate of time*)

think of this as the "continuous version" of the geometric distribution!

don't confuse this with the Poisson distribution just bc it's related with time, they're very different!
(Poisson is *number* of events in a certain time frame)

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

$$\mathrm{Var}(X) = \frac{1}{\lambda^2}$$

$$F_X(x) = 1 - e^{-\lambda x}$$

$F_X(x)$ = P(X <= x) *this is the integral of* $f_X(x)$

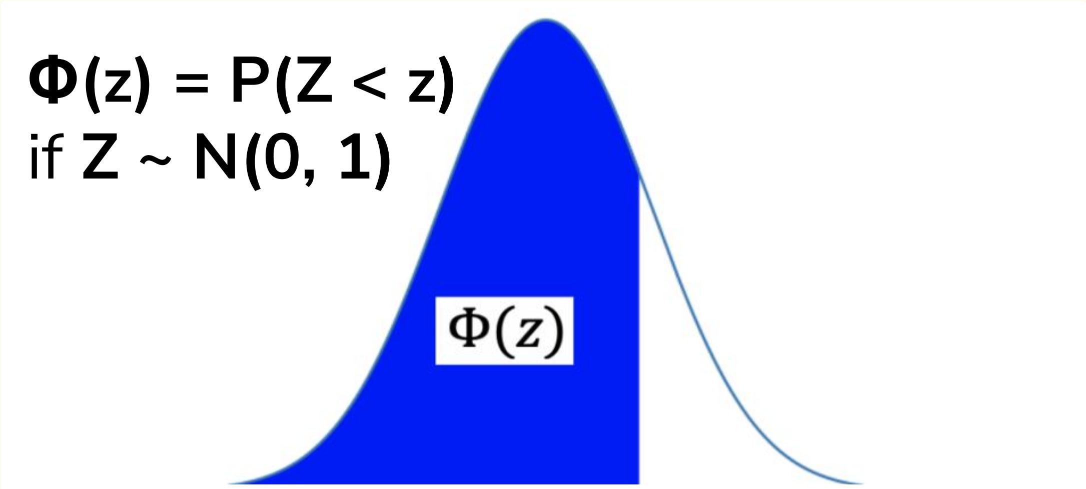
Review Normal Distribution

$X \sim N(\mu, \sigma^2)$
Standard normal is $Z \sim N(0, 1)$

general strategy:

1. **Find** μ and σ^2 of normal RV X
2. **Standardize** $X \sim N(\mu, \sigma^2)$ to get $Z = (X-\mu)/\sigma \sim N(0,1)$
3. Use **Phi table** to get appropriate value with $Z = (X-\mu)/\sigma$
4. **Solve** for X

PDF: $$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Review Normal Distribution

properties of the normal distribution

Closure for Normal Distribution

Let $X \sim N(\mu, \sigma^2)$. Then $aX + b \sim N(a\mu + b, b^2\sigma^2)$

"Reproductive" Property

Let X1,..,Xn be <u>independent</u> normal RVs with $E[X_i]=\mu_i$ and $Var(X_i=\sigma_i^2)$.

$$X = \sum_{i=1}^{n}(a_i X_i + b) \sim \mathcal{N}\left(\sum_{i=1}^{n}(a_i\mu_i + b), \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

Review CLT

Theorem. (Central Limit Theorem) $X_1, \ldots, X_n$ i.i.d. with mean μ and variance σ^2. Let $Y_n = \frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$. Then,

$$\lim_{n\to\infty} Y_n \to \mathcal{N}(0,1)$$

One main application:

Use Normal Distribution to Approximate Y_n

No need to understand Y_n !!

Example – Y_n is binomial

We understand binomial, so we can see how well approximation works

We flip n independent coins, heads with probability $p = 0.75$.

X = # heads $\quad \mu = \mathbb{E}(X) = 0.75n \quad \sigma^2 = \mathrm{Var}(X) = p(1-p)n = 0.1875n$

$\mathbb{P}(X \leq 0.7n)$

n	exact	$\mathcal{N}(\mu, \sigma^2)$ approx
10	0.4744072	0.357500327
20	0.38282735	0.302788308
50	0.25191886	0.207108089
100	0.14954105	0.124106539
200	0.06247223	0.051235217
1000	0.00019359	0.000130365

Example – Naive Approximation

Fair coin flipped (independently) **40** times. Probability of **20** or **21** heads?

Exact. $\mathbb{P}(X \in \{20,21\}) = \left[\binom{40}{20} + \binom{40}{21}\right]\left(\frac{1}{2}\right)^{40} \approx \boxed{0.2448}$

Approx. $X = \#$ heads $\quad \mu = \mathbb{E}(X) = 0.5n = 20 \quad \sigma^2 = \mathrm{Var}(X) = 0.25n = 10$

$$\mathbb{P}(20 \le X \le 21) = \Phi\left(\frac{20-20}{\sqrt{10}} \le \frac{X-20}{\sqrt{10}} \le \frac{21-20}{\sqrt{10}}\right)$$

$$\approx \Phi\left(0 \le \frac{X-20}{\sqrt{10}} \le 0.32\right)$$

$$= \Phi(0.32) - \Phi(0) \approx \boxed{0.1241}$$

😢

Example – Even Worse Approximation

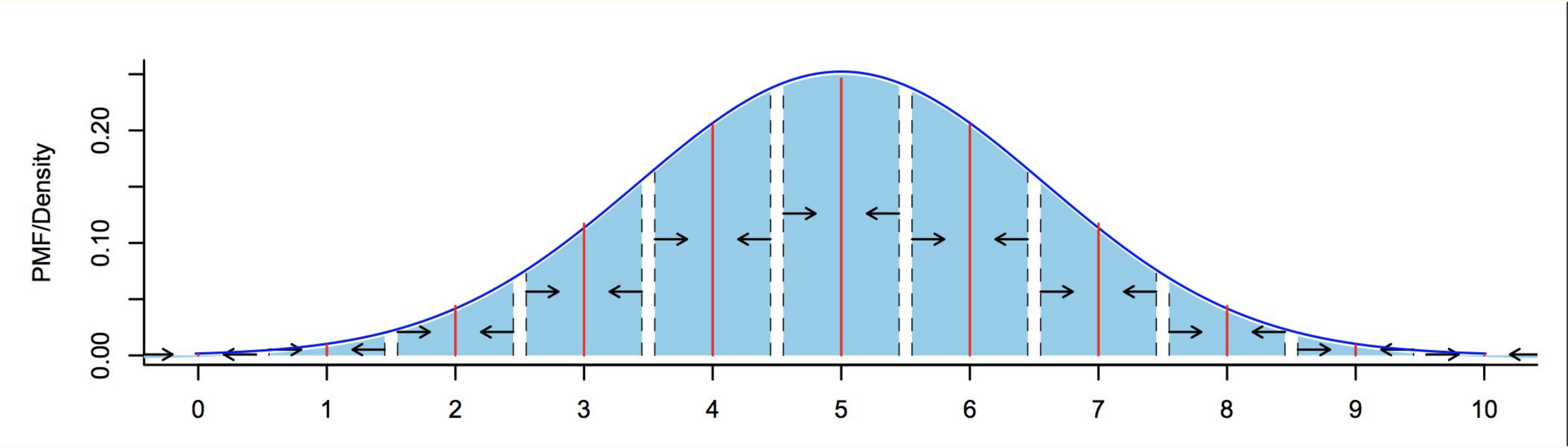
Fair coin flipped (independently) **40** times. Probability of **20** heads?

Exact. $\mathbb{P}(X = 20) = \binom{40}{20}\left(\frac{1}{2}\right)^{40} \approx \boxed{0.1254}$

Approx. $\mathbb{P}(20 \leq X \leq 20) = 0$ 😢

Solution – Continuity Correction

Probability estimate for i: Probability for all x that round to i!



To estimate probability that discrete RV lands in (integer) interval $\{a, \ldots, b\}$, compute probability continuous approximation lands in interval $[a - \frac{1}{2}, b + \frac{1}{2}]$

Example – Continuity Correction

Fair coin flipped (independently) **40** times. Probability of **20** or **21** heads?

Exact. $\mathbb{P}(X \in \{20,21\}) = \left[\binom{40}{20} + \binom{40}{21}\right]\left(\frac{1}{2}\right)^{40} \approx \boxed{0.2448}$

Approx. $X = \#\text{ heads}$ $\quad \mu = \mathbb{E}(X) = 0.5n = 20 \quad \sigma^2 = \mathrm{Var}(X) = 0.25n = 10$

$$\mathbb{P}(19.5 \le X \le 21.5) = \Phi\left(\frac{19.5 - 20}{\sqrt{10}} \le \frac{X - 20}{\sqrt{10}} \le \frac{21.5 - 20}{\sqrt{10}}\right)$$

$$\approx \Phi\left(-0.16 \le \frac{X - 20}{\sqrt{10}} \le 0.47\right)$$

$$= \Phi(0.47) - \Phi(-0.16) \approx \boxed{0.2452}$$

👍

Example – Continuity Correction

Fair coin flipped (independently) **40** times. Probability of **20** heads?

Exact. $\mathbb{P}(X = 20) = \binom{40}{20}\left(\frac{1}{2}\right)^{40} \approx \boxed{0.1254}$

Approx. $\mathbb{P}(19.5 \leq X \leq 20.5) = \Phi\left(\frac{19.5 - 20}{\sqrt{10}} \leq \frac{X - 20}{\sqrt{10}} \leq \frac{20.5 - 20}{\sqrt{10}}\right)$

$$\approx \Phi\left(-0.16 \leq \frac{X - 20}{\sqrt{10}} \leq 0.16\right)$$

$$= \Phi(0.16) - \Phi(-0.16) \approx \boxed{0.1272}$$

Task 3 – Batteries and exponential distributions (from Section 6)

Let X_1, X_2 be independent exponential random variables, where X_i has parameter λ_i, for $1 \leqslant i \leqslant 2$. Let $Y = \min(X_1, X_2)$.

a) Show that Y is an exponential random variable with parameter $\lambda = \lambda_1 + \lambda_2$. Hint: Start by computing $\mathbb{P}(Y > y)$. Two random variables with the same CDF have the same pdf. Why?

b) What is $\mathbb{P}(X_1 < X_2)$? (Use the continuous version of the law of total probability, conditioning on the probability that $X_1 = x$.)

c) You have a digital camera that requires two batteries to operate. You purchase n batteries, labelled $1, 2, \ldots, n$, each of which has a lifetime that is exponentially distributed with parameter λ, independently of all other batteries. Initially, you install batteries 1 and 2. Each time a battery fails, you replace it with the lowest-numbered unused battery. At the end of this process, you will be left with just one working battery. What is the expected total time until the end of the process? Justify your answer.

d) In the scenario of the previous part, what is the probability that battery i is the last remaining battery as a function of i? (You might want to use the memoryless property of the exponential distribution that has been discussed.)

Task 4 – Normal questions at the table (from Section 6)

a) Let X be a normal random with parameters $\mu = 10$ and $\sigma^2 = 36$. Compute $\mathbb{P}(4 < X < 16)$.

b) Let X be a normal random variable with mean 5. If $\mathbb{P}(X > 9) = 0.2$, approximately what is $\mathrm{Var}(X)$?

c) Let X be a normal random variable with mean 12 and variance 4. Find the value of c such that $\mathbb{P}(X > c) = 0.10$.

Task 6 – Tweets

A prolific Twitter user tweets approximately 350 tweets per week. Let's assume for simplicity that the tweets are independent, and each consists of a uniformly random number of characters between 10 and 140. (Note that this is a discrete uniform distribution.) Thus, the central limit theorem (CLT) implies that the number of characters tweeted by this user is approximately normal with an appropriate mean and variance. Assuming this normal approximation is correct, estimate the probability that this user tweets between 26,000 and 27,000 characters in a particular week. (This is a case where continuity correction will make virtually no difference in the answer, but you should still use it to get into the practice!).