PROPERTIES OF $E(X)$

- $E(a) = a$ FOR ANY CONSTANT a.
- FOR A LINEAR TRANSFORMATION $a + bX$,

$$E(a+bX) = a + bE(X).$$

NOTE: FOR A NONLINEAR TRANSFORMATION $g(X)$, $E[g(X)]$ USUALLY DIFFERS FROM $g(E(X))$, AS IN THE LAST EXAMPLE $E(X^2) \neq [E(X)]^2$.

Ex: FOR THE PREVIOUS EXAMPLE, FIND $E(7-2X)$.

A: BY DIRECT CALCULATION,

$$E(7-2X) = (7-0)0.4 + (7-2)0.3 + (7-4)0.2 + (7-6)0.1 = 5.$$

THE SMARTER WAY:

$$E(7-2X) = 7 - 2E(X) = 5.$$

MEAN AND VARIANCE

THE MEAN OF A DISCRETE RV X IS DEFINED AS

$$\mu = \mu_X = E(X).$$

FOR A LARGE SAMPLE OF OBSERVATIONS, WE EXPECT THE SAMPLE MEAN $\bar{X}$ TO BE CLOSE TO THE THEORETICAL MEAN μ.

RECALL THE SAMPLE VARIANCE IS THE AVERAGE OF SQUARED DISTANCES FROM THE SAMPLE MEAN:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

THE VARIANCE OF A RV X IS THE EXPECTED SQUARED DISTANCE FROM μ:

$$\sigma^2 = Var(X) = E[(X-\mu)^2].$$

A USEFUL ALTERNATIVE REPRESENTATION IS

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

EXERCISE: USE PROPERTIES OF $E(X)$ TO PROVE THAT

$$E[(X-\mu)^2] = E(X^2) - \mu^2.$$

THE STANDARD DEVIATION OF X IS THE POSITIVE SQUARE ROOT OF THE VARIANCE: $\sigma = \sqrt{Var(X)}$.

Ex: FIND THE VARIANCE FOR THE PREVIOUS EXAMPLE.

A: RECALL THAT $\mu = 1$ AND $E(X^2) = 2$, SO $\sigma^2 = E(X^2) - \mu^2 = 1$.

OR THE LONGER WAY:

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-1)^2 f(x)$$

$$= (0-1)^2 0.4 + (1-1)^2 0.3 + (2-1)^2 0.2 + (3-1)^2 0.1 = 1.$$

PROPERTIES OF VARIANCE

- $Var(X) \geq 0$, AND $Var(X) = 0 \iff X$ IS CONSTANT.
- $Var(X+a) = Var(X)$ FOR ANY CONSTANT a.
- $Var(bX) = b^2 Var(X)$ FOR ANY CONSTANT b.
- $\sigma_{a+bX} = |b| \sigma_X$.

IF A RV HAS LARGE VARIANCE, IT MEANS THAT OBSERVATIONS ARE EXPECTED TO VARY GREATLY.

Ex: FOR THE EXAMPLE $f(x)=0.1(4-x)$, WE FOUND $\sigma^2=1$.

- $Var\left(\frac{X}{2}\right)=\left(\frac{1}{2}\right)^2 Var(X)=\frac{1}{4}$

- $Var(X+6)=Var(X)=1$

- $Var(6-2X)=(-2)^2 Var(X)=4$

FOR A BINOMIAL DISTRIBUTION WITH n TRIALS AND p PROBABILITY OF SUCCESS, WE FIND THAT

$$\mu=E(X)=np$$

$$\sigma^2=Var(X)=np(1-p)$$

$$\sigma=\sqrt{Var(X)}=\sqrt{np(1-p)}$$

Ex: FIND THE MEAN AND STANDARD DEVIATION OF THE NUMBER X OF HEADS OBTAINED IN 100 TOSSES OF A COIN.

A: BINOMIAL DISTRIBUTION, $n=100$, $p=0.5$.

$$\mu=np=50$$

$$\sigma=\sqrt{np(1-p)}=5.$$

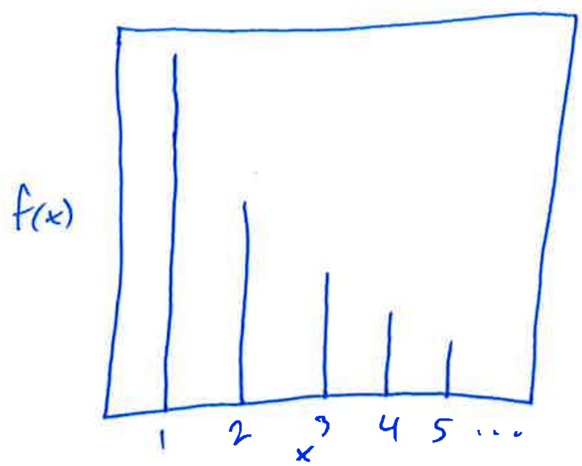
SO ALTHOUGH X WILL BE ABOUT 50 ON AVERAGE, IT WOULD NOT BE UNUSUAL TO OBSERVE VALUES BETWEEN 45 AND 55 ($\mu-\sigma$ AND $\mu+\sigma$).

GEOMETRIC DISTRIBUTION

THE GEOMETRIC DISTRIBUTION ARISES WHEN COUNTING THE NUMBER X OF TRIALS UNTIL THE FIRST SUCCESS. THE FINAL (SUCCESSFUL) TRIAL IS ALSO COUNTED, SO $X \in \{1, 2, \ldots, n\}$. THE DISCRETE PROBABILITY FUNCTION IS

$$f(x) = q^{x-1} p \ , \quad x = 1, 2, 3, \ldots, n.$$

THE SHAPE OF THE GRAPH IS STRONGLY SKEWED TO THE RIGHT.



TO FIND CDF OF $f(x) = q^{x-1}p$, WE GET

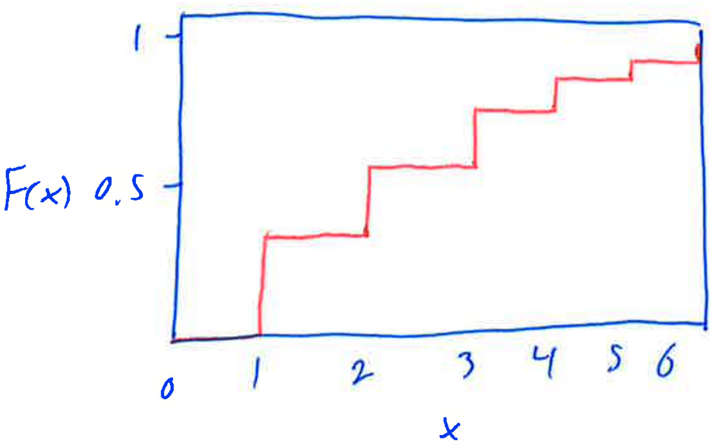
$$F(x) = p + qp + q^2 p + \ldots + q^{x-1} p$$
$$= p\,\frac{1-q^x}{1-q} = \boxed{1 - q^x}\ , \quad x = 1, 2, 3, \ldots, n.$$

IF CONVENIENT, YOU CAN ALSO USE $F(x) = P(X \le x) = 1 - P(X > x)$, WHERE $P(X > x)$ IS THE PROBABILITY OF NO SUCCESSES IN THE FIRST x TRIALS.

WE CAN CHECK THE CDF RESULT BY USING THE FORMULA $f(x) = F(x) - F(x-1)$:

$$f(x) = F(x) - F(x-1)$$
$$= 1 - q^x - (1 - q^{x-1})$$
$$= q^{x-1} - q^x$$
$$= q^{x-1}(1-q) = q^{x-1}p. \checkmark$$

THE CDF (GEOMETRIC OR OTHERWISE) APPROACHES 1 AS x INCREASES.



THE MEAN μ OF A GEOMETRIC RV X INVOLVES A SUMMATION TRICK THAT WE WILL NOT PROVE HERE.

$$E(X) = \sum_{x=1}^{\infty} x q^{x-1} p = p(1 + 2q + 3q^2 + \dots)$$
$$= p \frac{d}{dq}(1 + q + q^2 + q^3 + \dots)$$
$$= p \frac{d}{dq}\left(\frac{1}{1-q}\right) = \frac{p}{(1-q)^2}$$
$$= \boxed{\frac{1}{p}}$$

EX: A STUDENT GUESSES AT QUESTIONS WITH 5 MULTIPLE CHOICE ANSWERS EACH. LET X BE THE NUMBER OF THE FIRST QUESTION ANSWERED CORRECTLY. WHAT IS $E(X)$, AND WHAT IS THE PROBABILITY THAT $X > E(X)$?

A: X IS GEOMETRIC WITH $p = 0.2$, SO

$$E(X) = \frac{1}{p} = 5.$$

$$P(X>5) = 1 - P(X \leq 5) = 1 - F(5)$$

$$= 1 - \left[1 - (1-0.2)^5\right] = 0.3277.$$

SO ON AVERAGE, IT WILL TAKE THE STUDENT 5 QUESTIONS TO GET 1 RIGHT, AND THERE'S A 33% CHANCE THAT IT WILL TAKE MORE THAN 5 QUESTIONS.

POISSON DISTRIBUTION

POISSON IS A DISCRETE DISTRIBUTION THAT APPLIES WHEN WE COUNT THE NUMBER OF POINTS IN A GIVEN TIME/AREA/DISTANCE/VOLUME. FOR INSTANCE, THE NUMBER OF CARS THAT GO THROUGH AN INTERSECTION IN 10 MINUTES, OR THE NUMBER OF RUST SPOTS IN A $10\,m^2$ AREA.

LET λ BE THE AVERAGE RATE OF OCCURRENCES PER UNIT TIME (OR DISTANCE, AREA, VOLUME). THEN THE EXPECTED NUMBER μ OF OCCURRENCES IN AN INTERVAL OF LENGTH t IS

$$\mu = \lambda t.$$

FOR THE POISSON PROCESS, WE ASSUME

1) IF $N_{\Delta t}$ IS THE NUMBER OF OCCURRENCES IN A VERY SHORT TIME Δt, THEN

$$P(N_{\Delta t}=1) \simeq \lambda \Delta t,$$

$$P(N_{\Delta t}>1) \simeq 0.$$

2) COUNTS IN TIME INTERVALS THAT DON'T OVERLAP ARE INDEPENDENT.

TWO RVs X AND Y ARE <u>INDEPENDENT</u> IF

$$P(X \leq a, Y \leq b) = P(X \leq a) P(Y \leq b) \text{ FOR ALL } a, b.$$

i.e. PROBABILITIES INVOLVING ONE VARIABLE ARE UNAFFECTED BY INFORMATION ABOUT THE OTHER VARIABLE.

IF X AND Y ARE INDEPENDENT, THEN $E(XY) = E(X)E(Y)$.

THE PROCESS IS THE FOLLOWING.

- THE INTERVAL OF INTEREST $[0, t]$ IS SUBDIVIDED INTO n SUBINTERVALS OF LENGTH $\Delta t = t/n$.
- N_t IS APPROXIMATED BY COUNTING HOW MANY SUBINTERVALS CONTAIN AT LEAST ONE POINT.
- BY INDEPENDENCE OF SUBINTERVALS, N_t HAS AN APPROXIMATELY BINOMIAL DISTRIBUTION WITH $p = \lambda t/n$.
- THIS GIVES $\mu = np = \lambda t$, AND $\sigma^2 = np(1-p) = \lambda t(1 - \lambda t/n)$.
- AS $n \to \infty$, WE HAVE $\sigma^2 \to \lambda t$. SO $\sigma^2 = \mu$ AT THE LIMIT.
- LETTING $n \to \infty$, WE HAVE $P(N_t = x) = \binom{n}{x}\left(\frac{\lambda t}{n}\right)^x \left(1 - \frac{\lambda t}{n}\right)^{n-x} \to \frac{(\lambda t)^x}{x!} e^{-\lambda t}$.
- WITH $\mu = \lambda t$, WE OBTAIN THE <u>POISSON PROBABILITY FUNCTION</u>:

$$f(x) = \frac{\mu^x}{x!} e^{-\mu}, \quad x = 0, 1, 2, \ldots$$

THERE ARE INFINITELY MANY NONZERO PROBABILITIES, BUT THEY STILL ADD UP TO 1:

$$\sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} = e^{-\mu}\left(1+\mu+\frac{\mu^2}{2}+\frac{\mu^3}{6}+\ldots\right) = e^{-\mu}e^{\mu} = 1.$$

FOR A POISSON RV, $E(X) = Var(X) = \mu$.

EX: THE UOW SWITCHBOARD RECEIVES ON AVERAGE 0.6 CALLS PER MINUTE. FIND THE PROBABILITY THAT IN A 4-MINUTE INTERVAL THERE WILL BE (i) EXACTLY 3 CALLS, (ii) AT LEAST 3 CALLS.

A: THE RATE OF CALLS IS $\lambda = 0.6$ PER MINUTE, SO

$\mu = \lambda t = 0.6 \cdot 4 = 2.4$.

(i) $P(X=3) = \frac{2.4^3}{3!} e^{-2.4} \simeq 0.209$

R CODE:

```
dpois(3,2.4)
```

(ii) $P(X \geq 3) = 1 - P(X<3)$.

$P(X<3) = f(0) + f(1) + f(2)$

$$= \frac{2.4^0}{0!}e^{-2.4} + \frac{2.4^1}{1!}e^{-2.4} + \frac{2.4^2}{2!}e^{-2.4} = 0.5697$$

$\therefore P(X \geq 3) = 1 - 0.5697 = 0.4303$.

R CODE:

```
1-ppois(2,0.24)
```

EX: LET A = "AT LEAST 4 CALLS ARRIVE BETWEEN 10:00am AND 10:04am",

B = "EXACTLY 5 CALLS ARRIVE BETWEEN 10:00am AND 10:05am."

WHAT IS THE CONDITIONAL PROBABILITY OF A GIVEN B?

A: $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

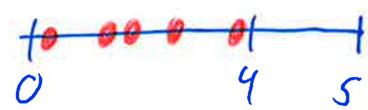
WE CAN FIND $P(B)$ DIRECTLY WITH $\mu = 0.6 \cdot 5 = 3$, $x = 5$.

$$P(B) = \frac{3^5}{5!} e^{-3}.$$

THERE ARE TWO POSSIBILITIES FOR $A \cap B$:

1) 4 CALLS IN $[0,4]$ AND 1 CALL IN $(4,5]$

2) 5 CALLS IN $[0,4]$ AND 0 CALLS IN $(4,5]$



SO WE NEED POISSON PROBABILITIES FOR EACH INTERVAL, WITH MEANS $0.6 \cdot 4 = 2.4$ FOR THE 4-MINUTE INTERVAL AND 0.6 FOR THE 1-MINUTE INTERVAL. BY INDEPENDENCE OF NON-OVERLAPPING COUNTS,

$$P(4 \text{ CALLS IN } [0,4] \text{ AND } 1 \text{ CALL IN } (4,5]) = \frac{2.4^4}{4!} e^{-2.4} \cdot \frac{0.6}{1!} e^{-0.6} = \frac{2.4^4 \cdot 0.6}{4!} e^{-3}$$

$$P(5 \text{ CALLS IN } [0,4] \text{ AND } 0 \text{ CALLS IN } (4,5]) = \frac{2.4^5}{5!} e^{-2.4} \frac{0.6}{0!} e^{-0.6} = \frac{2.4^5}{5!} e^{-3}$$

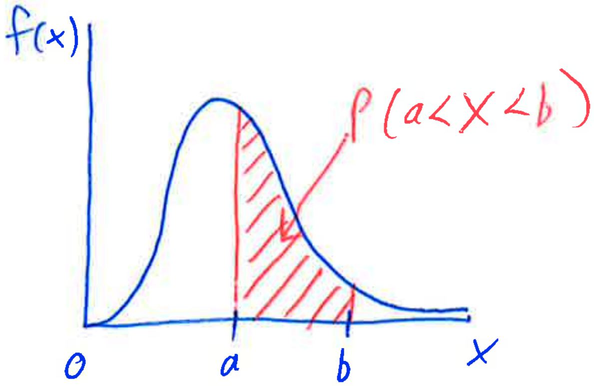
ADDING THESE TOGETHER, $P(A \cap B) = \frac{2.4^4 \cdot 0.6}{4!} e^{-3} + \frac{2.4^5}{5!} e^{-3}$

$$\therefore P(A|B) = \frac{2.4^4 \cdot 0.6/4! + 2.4^5/5!}{3^5/5!} = \boxed{0.7373}$$

CONTINUOUS RANDOM VARIABLES

A RV IS CONTINUOUS IF THE SET OF ITS POSSIBLE VALUES IS ONE OR MORE INTERVALS. eg. THE SET OF POSSIBLE VALUES T = TIME IT TAKES FOR A PIZZA TO ARRIVE TO YOUR HOUSE IS (HOPEFULLY) $\{t : 15 \le t \le 45 \text{ min}\}$.

THE PROBABILITY DENSITY FUNCTION (PDF) $f(x)$ OF A CONTINUOUS RV X IS THE FUNCTION SUCH THAT $P(a < X < b)$ IS THE AREA $\int_a^b f(x)\,dx$ UNDER THE CURVE $y = f(x)$ BETWEEN $x = a$ AND $x = b$.



NOTE: THE AREA IS THE SAME FOR $P(a \le X < b)$, $P(a < X \le b)$, $P(a \le X \le b)$.

A PDF CAN BE VIEWED AS A HISTOGRAM AS SAMPLE SIZE TENDS TO ∞, AND CLASS WIDTH TENDS TO ZERO. THE TOTAL AREA UNDER THE WHOLE CURVE MUST BE 1.

PROPERTIES OF PDF

- FOR VALUES OF x THAT ARE NEVER OBSERVED, $f(x)=0$.
- $f(x) \geq 0$ FOR ALL x.
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

EX: SHOW THAT $f(x) = 3x^2$; $x \in (0,1)$ IS A VALID PDF. FIND $P(0.2 < X < 0.9)$.

i) SINCE THE DOMAIN IS ONLY $(0,1)$, WE MAY SET $f(x)=0$ FOR ALL $x \notin (0,1)$.

ii) $3x^2 \geq 0$ FOR ALL x.

iii) $\int_{-\infty}^{\infty} 3x^2\,dx$? NO

$$\int_{-\infty}^{\infty} f(x)\,dx = \underbrace{\cancel{\int_{-\infty}^{0} f(x)\,dx}}_{0} + \int_{0}^{1} f(x)\,dx + \underbrace{\cancel{\int_{1}^{\infty} f(x)\,dx}}_{0} = \int_0^1 3x^2\,dx = x^3\Big|_0^1 = 1.$$

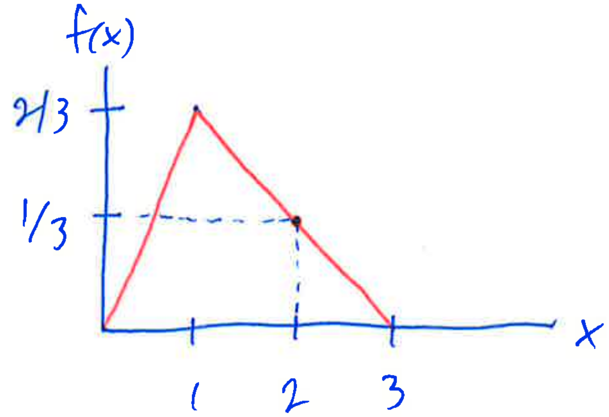
$$P(0.2 < X < 0.9) = \int_{0.2}^{0.9} 3x^2\,dx = 0.9^3 - 0.2^3 = 0.721.$$

THE CUMULATIVE DISTRIBUTION FUNCTION (CDF) OF A CONTINUOUS RV X IS DEFINED BY

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt.$$

IF YOU HAVE CDFs AVAILABLE, YOU CAN SOMETIMES AVOID INTEGRATION BY USING $P(a < X < b) = F(b) - F(a)$.

EX: LET $F(x)$ BE THE CDF OF THE CONTINUOUS RV WHOSE PROBABILITY FUNCTION $f(x)$ IS GRAPHED BELOW. FIND $F(2)$ AND $P(1 < X < 2)$.



A: TOTAL AREA UNDER f IS 1, SO THE AREA TO THE LEFT OF 2 IS

$$1 - (\text{RIGHT AREA}) = 1 - \frac{1/3}{2} = \frac{5}{6} = F(2).$$

FOR $P(1 < X < 2)$, EITHER FIND THE AREA BETWEEN 1 AND 2:

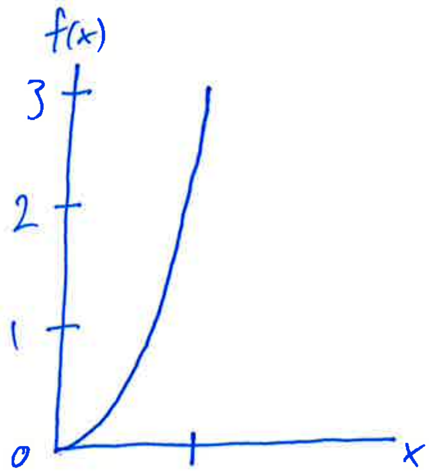
$$1 \cdot \frac{2/3 + 1/3}{2} = \frac{1}{2}$$

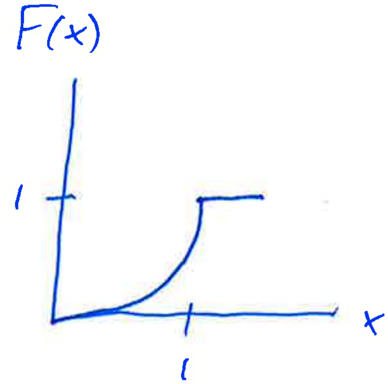
OR USE $F(2) - F(1) = \frac{5}{6} - \frac{1}{3} = \frac{1}{2}$.

PROPERTIES OF CDF

- AS $F(x)$ IS A PROBABILITY, IT MUST LIE BETWEEN 0 AND 1.
- AS x INCREASES, THE EVENT $\{X \leq x\}$ INCLUDES MORE OUTCOMES, SO $F(x)$ IS AN INCREASING FUNCTION OF x.
- FOR CONTINUOUS X, F IS CONTINUOUS. (FOR DISCRETE X, F IS A STEP FUNCTION).

EX: FOR THE PDF $f(x) = 3x^2$, $0 < x < 1$, F IS AS FOLLOWS.





WE OBTAIN F FROM f BY INTEGRATION, SO TO OBTAIN f FROM F, WE USE DIFFERENTIATION. BY THE FUNDAMENTAL THEOREM OF CALCULUS,

$$\frac{d}{dx}F(x) = \frac{d}{dx}\int_{-\infty}^{x} f(t)\,dt = f(x).$$

$$\text{i.e. } f(x) = \frac{d}{dx}F(x).$$

Ex: IF $F(x) = x^3$, $0 < x < 1$, THEN

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}x^3 = 3x^2, \quad 0 < x < 1.$$

MEAN, VARIANCE AND STANDARD DEVIATION

THE EXPECTED VALUE OF A FUNCTION OF A CONTINUOUS RV IS

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx.$$

THIS IS THE LONG-RUN AVERAGE VALUE OF $g(X)$ FOR A LARGE NUMBER OF EVALUATIONS. AS IN THE DISCRETE CASE, WE HAVE

- MEAN $\mu = E(X)$,
- VARIANCE $\sigma^2 = E[(X-\mu)^2] = E(X^2) - \mu^2$,
- STANDARD DEVIATION $\sigma = \sqrt{Var(X)}$.

Ex: FOR $f(x) = 3x^2$, $0 < X < 1$, FIND μ, $E(4X-2)$, $E\left(\frac{1}{X}\right)$, $\frac{1}{E(X)}$, $E(X^2)$, $Var(X)$ AND σ.

A: $\mu = E(X) = \int_0^1 x\, f(x)\, dx = \int_0^1 3x^3\, dx = \frac{3x^4}{4}\Big|_0^1 = \frac{3}{4}$

$E(4X-2) = 4E(X) - 2 = 1$

$E\left(\frac{1}{X}\right) = \int_0^1 \frac{1}{x} \cdot 3x^2\, dx = \frac{3x^2}{2}\Big|_0^1 = \frac{3}{2}$ (NOTE THAT $E\left(\frac{1}{X}\right) \neq \frac{1}{E(X)}$.)

$E(X^2) = \int_0^1 x^2 \cdot 3x^2\, dx = \frac{3x^5}{5}\Big|_0^1 = \frac{3}{5}$

$Var(X) = E(X^2) - \mu^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$

$\sigma = \sqrt{\frac{3}{80}} = \frac{1}{4}\sqrt{\frac{3}{5}}$.

MEDIAN

THE MEDIAN OF A CONTINUOUS RV X IS FOUND BY SOLVING $P(X \leq Q_2) = F(Q_2) = 0.5$. WE WANT HALF THE AREA UNDER THE PDF TO BE ON EITHER SIDE OF $x = Q_2$. SIMILARLY, THE UPPER AND LOWER QUARTILES SATISFY $F(Q_3) = 0.75$, $F(Q_1) = 0.25$.

IF f IS SYMMETRIC ABOUT μ, THEN $Q_2 = \mu$.

EX: LET $F(x) = x^3$, $0 < x < 1$. FIND μ AND Q_2. WHAT DOES THIS SAY ABOUT f?

A: $f(x) = \frac{d}{dx} F(x) = 3x^2$. FROM THE PREVIOUS EXAMPLE,

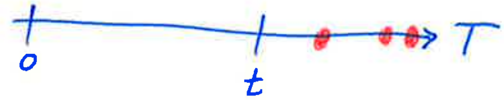
$\mu = \frac{3}{4}$.

$F(Q_2) = 0.5 \Rightarrow Q_2^3 = 0.5 \Rightarrow Q_2 = 0.5^{1/3} \approx 0.7937$.

THE MEDIAN IS BIGGER THAN THE MEAN, SO f IS SKEWED TO THE LEFT.

EXPONENTIAL DISTRIBUTION

THE EXPONENTIAL DISTRIBUTION CAN BE GENERATED FROM THE POISSON PROCESS. LET T BE THE WAITING TIME UNTIL THE 1^{st} POINT WITH RATE λ PER UNIT TIME. THEN $P(T > t) = P(\text{NO POINTS IN } [0,t])$.



THE NUMBER OF POINTS IN $[0,t]$ IS A POISSON RV WITH MEAN λt:

$$P(T > t) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

$$F(t) = P(T \le t) = 1 - e^{-\lambda t}$$

$$f(t) = \frac{d}{dt} F(t) = \lambda e^{-\lambda t}, \quad t \ge 0$$

<u>DEF</u>: T HAS AN <u>EXPONENTIAL DISTRIBUTION</u> WITH RATE PARAMETER λ

IF THE PDF AND CDF HAVE THE FORM

$$f(t) = \lambda e^{-\lambda t}, \quad t \ge 0,$$

$$F(t) = 1 - e^{-\lambda t}, \quad t \ge 0.$$

THE APPLICATIONS (SOMETIMES SEEN WITH PARAMETER $\beta = 1/\lambda$) ARE FOR INTERARRIVAL TIMES OF CUSTOMERS IN QUEUE, AND TIME UNTIL FAILURE IN RELIABILITY MODELS.

EXPONENTIAL MEAN AND MEDIAN

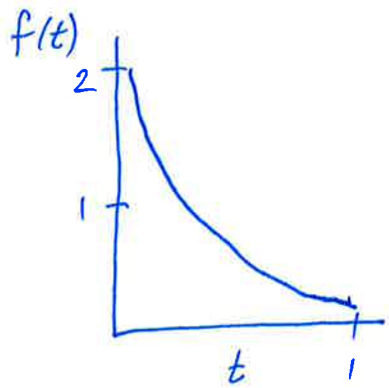
AN EXPONENTIAL RV HAS $\mu = \frac{1}{\lambda}$ AND $\sigma = \frac{1}{\lambda}$.

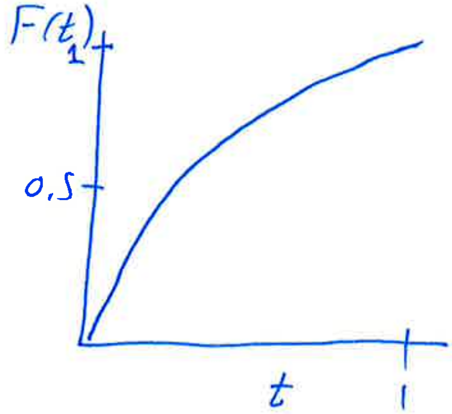
<u>EXERCISE</u>: PROVE IT! (HINT: INTEGRATE BY PARTS)

TO FIND THE MEDIAN, SOLVE $F(Q_2) = 1 - e^{-\lambda Q_2} = \frac{1}{2}$

$$-\lambda Q_2 = \ln \frac{1}{2} \Rightarrow \boxed{Q_2 = \frac{1}{\lambda} \ln 2.}$$

EX: THE EXPONENTIAL PDF AND CDF GRAPHS FOR $\lambda = 2$:





EX: THE UOW SWITCHBOARD RECEIVES ON AVERAGE 0.6 CALLS/min. ACCORDING TO A POISSON PROCESS. THE FIRST CALL OF THE DAY ARRIVES AT T MINUTES AFTER 9am.

(i) WHAT IS THE PROBABILITY THAT $T < 2$ min.?

(ii) WHAT IS THE MEDIAN OF T?

A: (i) SINCE T IS AN EXPONENTIAL RV WITH $\lambda = 0.6$, WE HAVE

$$F(t) = 1 - e^{-0.6t}.$$

$$P(T < 2) = F(2) = 1 - e^{-0.6 \cdot 2} \approx 0.6988.$$

R CODE:

```
pexp(2,0.6)
```

(ii) $F(Q_2) = \frac{1}{2} = 1 - e^{-0.6 Q_2} \Rightarrow Q_2 = \frac{\ln 2}{0.6} \approx 1.155.$

NOTE THAT THE MEDIAN IS LESS THAN THE EXPONENTIAL MEAN $\frac{1}{\lambda} = \frac{5}{3}$, SO f IS SKEWED TO THE RIGHT.

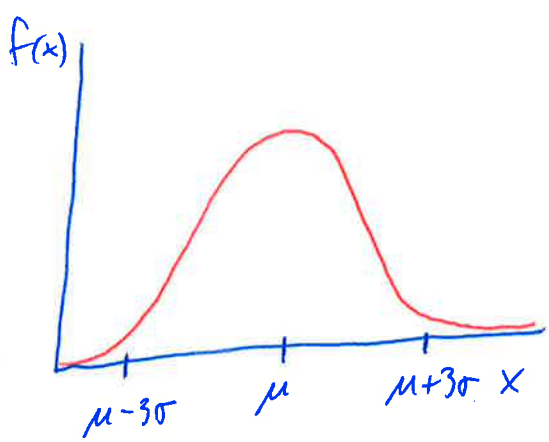
NORMAL DISTRIBUTION

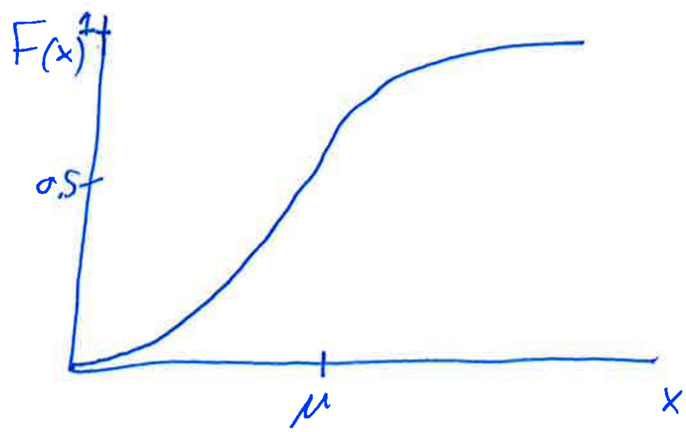
A NORMAL (GAUSSIAN) RV HAS PDF

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

THE NOTATION $N(\mu, \sigma^2)$ IS OFTEN USED. FOR A STANDARD NORMAL RV, $\mu = 0$ AND $\sigma = 1$.

- THE NORMAL PDF IS SYMMETRIC, BELL-SHAPED AND CHARACTERIZED BY μ AND σ. THERE ARE NO UPPER/LOWER BOUNDS, BUT IT IS VERY UNLIKELY TO OBSERVE VALUES MORE THAN 3σ AWAY FROM THE MEAN.





ANALYTIC EXPRESSIONS DON'T EXIST FOR THE NORMAL PDF, SO INTEGRATION ISN'T POSSIBLE. PROBABILITIES ARE FOUND NUMERICALLY, BY TABLES OR SOFTWARE.

R CODE:

pnorm

ANY NORMAL RV CAN BE STANDARDIZED WITH THE VARIABLE CHANGE

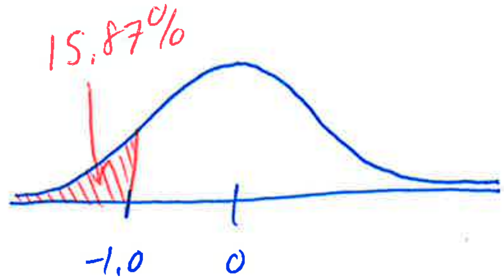
$$Z = \frac{X-\mu}{\sigma}, \text{ THEN}$$

$$E(Z) = \frac{E(X)-\mu}{\sigma} = 0 \text{ AND } Var(Z) = \frac{\sigma^2}{\sigma^2} = 1.$$

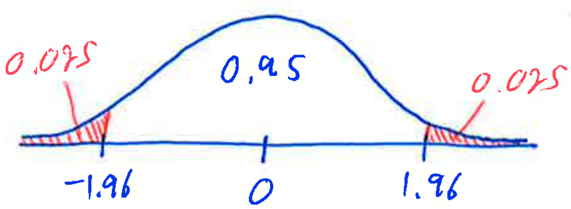
SO FOR $X \sim N(\mu, \sigma^2)$, $Z \sim N(1, 0)$. THEREFORE, STANDARD NORMAL TABLES ARE SUFFICIENT FOR ANY NORMAL PROBLEM:

$$P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right)$$

Z	.00	.03	.06	.09 ← (2nd DECIMAL)
-1.9	.0287	.0268	.0250	.0233
-1.0	.1587	.1515	.1446	.1379
0.0	.5000	.5120	.5239	.5359
1.9	.9713	.9732	.9750	.9767



DIFFERENT BOOKS/WEBSITES HAVE SLIGHTLY DIFFERENT FORMAT. IT MAY BE LEFT-HAND OR RIGHT-HAND AREAS, OR AREA BETWEEN Z AND 0. AREAS FOR Z < 0 ARE OFTEN EXCLUDED, AS THEY CAN BE FOUND BY SYMMETRY.

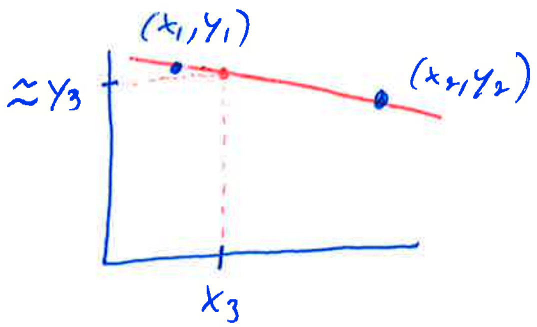


$$P(Z < -1.96) = P(Z > 1.96) = 0.025$$

LINEAR INTERPOLATION CAN BE USED TO GET MORE ACCURACY.

INTERPOLATION CONNECTS DISCRETE POINTS WITH STRAIGHT LINES, WHICH ALLOWS ESTIMATION OF VALUES THAT ARE UNAVAILABLE. FOR EXAMPLE, IF YOU HAVE POINTS (x_1, y_1) AND (x_2, y_2) ONLY, BUT YOU WANT TO KNOW APPROXIMATELY WHAT THE Y-VALUE OF x_3 IS WHEN x_3 IS $\frac{1}{5}$ OF THE WAY BETWEEN x_1 AND x_2, THEN

$$y_3 \approx y_1 + \frac{1}{5}(y_2 - y_1):$$



TO FIND PERCENTAGE/PROPORTION/PROBABILITY/AREA:

1) CALCULATE $z = \frac{x - \mu}{\sigma}$ FOR THE ENDPOINT(S).
2) READ THE AREA(S) FROM A Z-TABLE.
3) $P(Z < z)$ IS DIRECT FROM THE TABLE.
4) $P(Z > z) = 1 - P(Z < z)$.

$$P(z_1 \leq Z \leq z_2) = P(Z \leq z_2) - P(Z \leq z_1)$$

EX: ROLLS OF ROOF CLADDING HAVE NORMALLY DISTRIBUTED WEIGHTS WITH $\mu = 42$ kg AND $\sigma = 4.4$ kg. WHAT PROPORTION OF THESE ROLLS WEIGH

i) MORE THAN 44 kg?

ii) BETWEEN 39 AND 44 kg?

A: LET X BE THE WEIGHT OF A RANDOMLY CHOSEN ROLL. THEN $X \sim N(42, 4.4^2)$.

i) $P(X \leq 44) = P\left(Z \leq \frac{44-42}{4.4}\right) = P(Z \leq 0.4545)$

LOOK UP THE NORMAL TABLES FOR $Z = 0.45$ AND $Z = 0.46$, THEY ARE 0.6736 AND 0.6772, RESPECTIVELY. THEN LINEARLY INTERPOLATE TO APPROXIMATE $P(Z \leq 0.4545)$:

$$P(X \leq 44) \approx 0.6736 + 0.45(0.6772 - 0.6736)$$
$$= 0.6752$$

$$P(X > 44) \approx 1 - 0.6752 = \boxed{0.3248}$$

ii) $P(39 \leq X \leq 44) = P\left(\frac{39-42}{4.4} \leq Z \leq \frac{44-42}{4.4}\right)$

$$= P(-0.6818 \leq Z \leq 0.4545)$$
$$= P(Z \leq 0.4545) - P(Z < -0.6818)$$

WE ALREADY HAVE $P(Z \leq 0.4545) = 0.6752$.

THE QUICK METHOD: ROUND -0.6818 AND LOOK UP $Z = -0.68$.

$P(Z < -0.68) = 0.2483$.

THE MORE ACCURATE METHOD: INTERPOLATE USING $Z = -0.69$

(MAKE SURE YOU GO THE RIGHT DIRECTION WITH NEGATIVE NUMBERS!)

$$P(Z < -0.6818) \approx 0.2483 + 0.18(0.2451 - 0.2483) = 0.2477$$

$$P(39 \leq X \leq 44) = 0.6752 - 0.2477 = \boxed{0.4275}$$