1)

1. Raindrops are falling at an average rate of 20 drops per square inch per minute. What would be a reasonable distribution to use for the number of raindrops hitting a particular region measuring 5 inches^2 in t minutes? Why? Using your chosen distribution, compute the probability that the region has no rain drops in a given 3 second time interval.

Poisson: $\text{Pois}(20 \cdot 5 \cdot t)$
$= \text{Pois}(100t)$

Poisson distribution can be used to find a distribution given the rate of occurrence of events.

Since rainfall / inch^2 / min is 20, then rainfall in 5 inch^2 and t min is
$20 \cdot 5 \cdot t = 100t$

Let R be r.v for no. of raindrops in 5 inches^2 over 3 seconds.

$$R \sim \text{Pois}(20 \cdot 5 \cdot \tfrac{3}{60})$$
$$= \text{Pois}(5)$$

$$P(R=k) = e^{-5} \frac{5^k}{k!}$$

$$P(R=0) = e^{-5} \cdot \frac{5^0}{0!} = e^{-5} \quad //$$

2.

2. Harvard Law School courses often have assigned seating to facilitate the "Socratic method." Suppose that there are 100 first year Harvard Law students, and each takes two courses: Torts and Contracts. Both are held in the same lecture hall (which has 100 seats), and the seating is uniformly random and independent for the two courses.

(a) Find the probability that no one has the same seat for both courses (exactly; you should leave your answer as a sum).

(b) Find a simple but accurate approximation to the probability that no one has the same seat for both courses.

(c) Find a simple but accurate approximation to the probability that at least two students have the same seat for both courses.

a) Let $S_1, \ldots, S_{100}$ be events s.t S_i is event where person i is in the same seat.

$$P(S_1^c \cap S_2^c \cap \ldots \cap S_{100}^c) = 1 - P(S_1 \cup S_2 \cup \ldots \cup S_{100})$$

$$P(S_1 \cup S_2 \cup \ldots \cup S_{100}) = \sum_{i=1}^{100} P(S_i) - \sum_{i<j}^{100} P(S_i \cap S_j) + \ldots - P(S_1 \cap S_2 \cap \ldots \cap S_{100})$$

$$= 100 P(S_i) - \binom{100}{2} P(S_i \cap S_j) + \binom{100}{3} P(S_0 \cap S_j \cap S_k) - \ldots$$
$$+ P(S_1 \cap S_2 \cap \ldots \cap S_{100})$$

$$= 100 \cdot \frac{1}{100} - \binom{100}{2} \frac{1}{100} \cdot \frac{1}{99} + \binom{100}{3} \frac{1}{100} \cdot \frac{1}{99} \cdot \frac{1}{98} - \binom{100}{4} \frac{1}{100} \cdot \frac{1}{99} \cdot \frac{1}{98} \cdot \frac{1}{97}$$
$$+ \ldots$$

$$= \sum_{k=1}^{100} (-1)^{k+1} \binom{100}{k} \frac{(100-k)!}{100!}$$

$$\Rightarrow P(S_1^c \cap S_2^c \cap \ldots \cap S_{100}^c) = 1 - \sum_{k=1}^{100} (-1)^{k+1} \binom{100}{k} \frac{(100-k)!}{100!}$$

$$= 1 - \sum_{k=1}^{100} (-1)^{k+1} \frac{100!}{k!(100-k)!} \cdot \frac{(100-k)!}{100!}$$

$$= 1 - \sum (-1)^{k+1} \frac{1}{k!}$$

$$= 1 - \sum \frac{(-1)^{k+1}}{k!}$$

$$= 1 - \sum_{k=1}^{100} (-1)^{k+1} \frac{1}{k!} \qquad e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}$$

$$= 1 - \left(1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots\right)$$

$$= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots$$

$$\approx e^{-1}$$

c) Let X be a r.v for the no. of students with the same seats.

$$E(X) = 100 \cdot P(S_i)$$
$$= 100 \cdot \frac{1}{100} = 1$$

$$X \sim Pois(1)$$

$$P(X = k) = e^{-1} \frac{1}{k!}$$

$$P(X \geq 2) = 1 - P(X < 2)$$
$$= 1 - P(X=1) - P(X=0)$$
$$= 1 - e^{-1} - e^{-1}$$
$$= 1 - 2e^{-1}$$

3)

3. Let X be a Pois(λ) random variable, where λ is fixed but unknown. Let $\theta = e^{-3\lambda}$, and suppose that we are interested in estimating θ based on the data. Since X is what we observe, our estimator is a function of X, call it $g(X)$. The *bias* of the estimator $g(X)$ is defined to be $E(g(X)) - \theta$, i.e., how far off the estimate is on average; the estimator is *unbiased* if its bias is 0.

(a) For estimating λ, the r.v. X itself is an unbiased estimator. Compute the bias of the estimator $T = e^{-3X}$. Is it unbiased for estimating θ?

(b) Show that $g(X) = (-2)^X$ is an unbiased estimator for θ. (In fact, it is the best unbiased estimator, in the sense of minimizing the average squared error.)

a) Bias of T: $E(T) - \theta$

$= E(e^{-3X}) - \theta$

Since X is unbiased, unbiased for λ not θ

$E(X) - \theta = 0$

$\lambda - \theta = 0$

$\lambda = \theta$

$\cancel{\lambda = e^{-3\lambda}}$

$E(e^{-3X}) \Rightarrow$ Let $f(x) = e^{-3x}$

$E(f(X)) = E(T)$

$$\cancel{\begin{aligned} E(f(x)) &= \sum_{x \in X} f(x) P(X=x) \\ &= \sum_{x \in X} e^{-3x} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \sum_{x \in X} e^{-3x-\lambda} \frac{\lambda^x}{x!} \\ &= \sum_{x \in X} e^{-3x-\lambda} \cdot e^{-3x\lambda} \cdot \frac{1}{x!} \\ &= \sum_{x \in X} e^{-3x(1+\lambda)-\lambda} \frac{1}{x!} \end{aligned}}$$

$$\begin{aligned} E(T) - \theta &= \left[\sum_{x \in X} e^{-3x(1+\lambda)-\lambda} \frac{1}{x!}\right] - e^{-3\lambda} \\ &= \left[e^{-\lambda} \sum_{x \in X} e^{-3x(1+\lambda)} \frac{1}{x!}\right] - e^{-3\lambda} \\ &= e^{-\lambda}\left[\left(\sum_{x \in X} e^{-3x(1+\lambda)} \frac{1}{x!}\right) - e^{-2\lambda}\right] \\ &= e^{-\lambda}\left[\left(\sum_{x=0}^{\infty} \frac{[e^{-3(1+\lambda)}]^x}{x!}\right) - e^{-2\lambda}\right] \\ &= e^{-\lambda}\left[e^{e^{-3(1+\lambda)}} - e^{-2\lambda}\right] \neq 0 \Rightarrow \text{biased.} \end{aligned}$$

$\frac{k^x}{x!}$

b) $g(X) = (-2)^X$ is unbiased.

Pf:
$$\begin{aligned} E(g(X)) &= \sum_{x=0}^{\infty} (-2)^x P(X=x) \\ &= \sum_{x=0}^{\infty} (-2)^x e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \sum_{x=0}^{\infty} e^{-\lambda} \frac{(-2\lambda)^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(-2\lambda)^x}{x!} = e^{-\lambda} \cdot e^{-2\lambda} \\ &= e^{-3\lambda} \end{aligned}$$

$-2^2 \cdot 3^3$

$\Rightarrow E(g(X)) - \theta = e^{-3\lambda} - e^{-3\lambda} = 0 \Rightarrow$ unbiased //

(c) Explain intuitively why $g(X)$ is a silly choice for estimating θ, despite (b), and show how to improve it by finding an estimator $h(X)$ for θ that is always at least as good as $g(X)$ and sometimes strictly better than $g(X)$. That is,

$$|h(X) - \theta| \leq |g(X) - \theta|,$$

with the inequality sometimes strict.

Increases variance of the estimator significantly.

$g(X)$ can be negative

$$g(x) = (-2)^x$$

$$\begin{aligned} Var(g(x) - \theta) &= Var(g(x)) \\ &= E([g(x)]^2) - E(g(x))^2 \\ &= E((-2)^{2x}) - e^{-6\lambda} \end{aligned}$$

$$\begin{aligned} E((-2)^{2x}) &= \sum_{x=0}^{\infty} (-2)^{2x} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \sum_{x=0}^{\infty} 4^x e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=0}^{\infty} e^{-\lambda} \frac{(4\lambda)^x}{x!} \\ &= \sum_{x=0}^{\infty} e^{-\lambda} e^{4\lambda} \\ &= e^{3\lambda} \end{aligned}$$

$$\Rightarrow Var(g(x) - \theta) = e^{3\lambda} - e^{-6\lambda}$$

Let $h(x) = x$

$$\begin{aligned} Var(X - \theta) &= Var(X) \\ &= \lambda \end{aligned}$$

$$\lambda < e^{3\lambda} - e^{-6\lambda} \text{ For } \lambda \geq 0.$$

$\Rightarrow$ X is a better estimator than $g(x)$.

$\Box$.

1. Let $Z \sim \mathcal{N}(0,1)$ and let S be a "random sign" independent of Z, i.e., S is 1 with probability 1/2 and -1 with probability 1/2. Show that $SZ \sim \mathcal{N}(0,1)$.

For $SZ \sim N(0,1)$,

$$P(SZ \leq k) = \Phi(k)$$

$$\begin{aligned}
P(SZ \leq k) &= P(-Z \leq k)\cdot\tfrac{1}{2} + P(Z \leq k)\cdot\tfrac{1}{2} \quad \text{By LOTP} \\
&= P(Z \geq -k)\cdot\tfrac{1}{2} + P(Z \leq k)\cdot\tfrac{1}{2} \\
&= \tfrac{1}{2}[1 - P(Z \leq -k) + \Phi(k)] \qquad Z \geq -k \\
&= \tfrac{1}{2}[1 - (1 - P(Z \leq k)) + \Phi(k)] \\
&= \tfrac{1}{2}[1 - 1 + \Phi(k) + \Phi(k)] \\
&= \tfrac{1}{2}[2\Phi(k)] = \Phi(k)
\end{aligned}$$

$\Rightarrow P(SZ \leq k) = \Phi(k)$

$\Rightarrow SZ \sim N(0,1)$

$\square$.

2) 2. Explain why $P(X < Y) = P(Y < X)$ if X and Y are i.i.d. Does it follow that $P(X < Y) = 1/2$? Is it still always true that $P(X < Y) = P(Y < X)$ if X and Y have the same distribution but are not independent?

Suppose X, Y are i.i.d.

X and Y have same distribution and are independent. Hence.

$P(X<Y) = P(Y<X)$ by symmetry.

$$\begin{aligned}
P(X<Y) &= 1 - P(X \geq Y) \\
&= 1 - P(Y \leq X)
\end{aligned}$$

If $P(X<Y) = P(Y<X)$ and X, Y are continuous, then

$P(Y \leq X) = P(Y<X)$

$\Rightarrow P(X<Y) = 1 - P(Y<X)$

$P(X<Y) = \frac{1}{2}$ $\square$

But if X, Y are not continuous, then $P(X<Y) \neq \frac{1}{2}$.

Not true.

3. Explain why if $X \sim \mathrm{Bin}(n, p)$, then $n - X \sim \mathrm{Bin}(n, 1-p)$.

$$X \sim \mathrm{Bin}(n,p) \Rightarrow P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$n-X \Rightarrow P(n-X=k) = P(X=n-k)$$
$$= \binom{n}{n-k} p^{n-k} (1-p)^k$$
$$= \binom{n}{k} (1-p)^k p^{n-k}$$
$$\Rightarrow n-X \sim \mathrm{Bin}(n, 1-p)$$

$\square$.

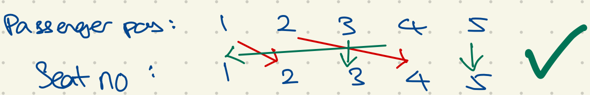
4) 4. There are 100 passengers lined up to board an airplane with 100 seats (with each seat assigned to one of the passengers). The first passenger in line crazily decides to sit in a randomly chosen seat (with all seats equally likely). Each subsequent passenger takes his or her assigned seat if available, and otherwise sits in a random available seat. What is the probability that the last passenger in line gets to sit in his or her assigned seat? (This is another common interview problem, and a beautiful example of the power of symmetry.)

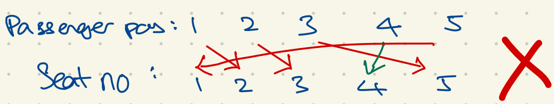
Hint: call the seat assigned to the jth passenger in line "Seat j" (regardless of whether the airline calls it seat 23A or whatever). What are the possibilities for which seats are available to the last passenger in line, and what is the probability of each of these possibilities?

If last passenger doesn't get to sit in their seat, then they must sit in the crazy guy's seat.

For seating error to not be propogated, next person with taken seat must sit in crazy guy's seat.

<u>Example</u>: suppose queue position corresponds to seat no. e.g 3rd person in queue sits at no. 3.





We define the 'correct' seat to be the seat that does not propogate the seating error further e.g person sits in assigned seat or crazy guy seat.

Let X_i be the event that in a flight with i seats and i people, the first person takes a random seat.

Let F be the event that the last person does not get his seat.
Let C be the event that the first person takes the correct seat.

$$P(F \mid X_n) = P(F \mid X_n \cap C) \cdot P(C) + P(F \mid X_n \cap C^c) P(C^c)$$
$$= 0 + P(F \mid X_n \cap C^c) \frac{n-1}{n}$$

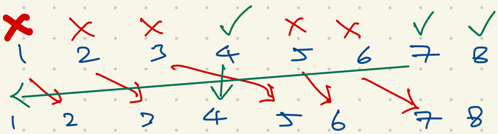
$$P(F|X_n \cap C^c)\,\frac{n-1}{n} = \frac{n-1}{n}\left[P(F|X_{n-1})\cdot\frac{1}{n-1} + P(F|X_{n-2})\cdot\frac{1}{n-1} + \dots + P(F|X_2)\cdot\frac{1}{n-1} + \frac{1}{n-1}\right]$$
$$= \frac{1}{n}\left[P(F|X_{n-1}) + P(F|X_{n-2}) + \dots + 1\right]$$
$$= \frac{1}{n}\left[1 + P(F|X_2) + P(F|X_3) + \dots + P(F|X_{n-1})\right]$$
$$= \frac{1}{n}\left[1 + \tfrac{1}{2}(P(F|X_1)) + \tfrac{1}{3}(P(F|X_2) + P(F|X_1)) + \tfrac{1}{4}(P(F|X_3) + \dots\right.$$

= nevermind, this is a mess.

Total possible ways to propogate error: 2^{n-1}.

Since the error is a chain of people that didn't get their assigned seat, this is equivalent to the total no. of combinations of $n-1$ people.

e.g suppose we pick a combination of people: 2, 3, 5, 6 out of 7 people. This is equivalent to:



The no. of error chains that end with a correct last seating is 2^{n-2}.

Since probability of a specific seating arrangement are all identical,

$$P(F) = \frac{2^{n-2}}{2^{n-1}} = \frac{1}{2}$$

1. Let $Y = e^X$, where $X \sim \mathcal{N}(\mu, \sigma^2)$. Then Y is said to have a *LogNormal* distribution; this distribution is of great importance in economics, finance, and elsewhere. Find the CDF and PDF of Y (the CDF should be in terms of Φ).

$Y = e^X$ where $X \sim N(\mu, \sigma^2)$

CDF of X: $P(X \le k) = \sigma\Phi(k) + \mu$

CDF of Y:
$$P(Y \le y) = P(e^X \le y)$$
$$= P(X \geq \ln y)$$
$$= P\left(\frac{X-\mu}{\sigma} \geq \frac{\ln y - \mu}{\sigma}\right)$$
$$= \Phi\left(\frac{\ln y - \mu}{\sigma}\right)$$

PDF of Y: $\varphi\left(\frac{\ln y - \mu}{\sigma}\right)\cdot\frac{1}{y\sigma}$

$$= \frac{1}{\sqrt{2\pi}}\, e^{-\left(\frac{\ln y - \mu}{\sigma}\right)^2/2}\cdot\frac{1}{y\sigma}$$

2. Let $U \sim \text{Unif}(0,1)$. Using U, construct a r.v. X whose PDF is $\lambda e^{-\lambda x}$ for $x > 0$ (and 0 otherwise), where $\lambda > 0$ is a constant. Then X is said to have a *Exponential* distribution; this distribution is of great importance in engineering, chemistry, survival analysis, and elsewhere.

$U \sim \text{Unif}(0,1)$

$X_{PDF} = \lambda e^{-\lambda x}$

$$X_{CDF} = \int_0^t \lambda e^{-\lambda x}\,dx = \lambda \int_0^t e^{-\lambda x}\,dx$$
$$= -\left[e^{-\lambda x}\right]_0^t$$
$$= -e^{-\lambda t} + e^0$$
$$= -e^{-\lambda t} + 1$$

$$P(X \leq t) = F(t) = -e^{-\lambda t} + 1$$

$$F^{-1}(t): \quad y = -e^{-\lambda x} + 1$$
$$1 - y = e^{-\lambda x}$$
$$-\lambda x = \ln(1-y)$$
$$x = -\frac{\ln(1-y)}{\lambda}$$

$$\Rightarrow F^{-1}(t) = -\frac{\ln(1-y)}{\lambda}$$

$$\Rightarrow F(t) = -\frac{\ln(1-U)}{\lambda} \quad //$$

Let $Z \sim \mathcal{N}(0,1)$. Find $E(\Phi(Z))$ *without* using LOTUS, where Φ is the CDF of Z.

$Z \sim N(0,1)$

Let $X = \Phi(Z)$

By universality of the uniform,

$\Phi(Z) = U$ since Φ is CDF of $Z \sim N(0,1)$.

$\Phi(Z) = U$

$$\Rightarrow E(\Phi(Z)) = \int_0^1 x\,dx = \frac{1}{2}x^2\Big|_0^1 = \frac{1}{2} \quad //$$

A stick is broken into two pieces, at a uniformly random chosen break point. Find the CDF and average of the length of the longer piece.

Let ℓ be the length of the stick.

Let L be a r.v for the length of the longer piece.
Let B be a r.v for the breakpoint, where $0 \leq B \leq \frac{\ell}{2}$

$B \sim U$ on interval $(0, \frac{\ell}{2})$

$L = \ell - B$

CDF of L: $P(L \leq k) = P(\ell - B \leq k)$

$$= P(-B \leq k - \ell)$$
$$= P(B \geq \ell - k)$$
$$= 1 - P(B \leq \ell - k)$$
$$= 1 - \frac{\ell - k}{(\ell/2)}$$
$$= 1 - \frac{2(\ell - k)}{\ell}$$
$$= 1 - 2 + \frac{2k}{\ell} = 2\frac{k}{\ell} - 1 \quad //$$

Average length: $\frac{3}{4}\ell$ //

1. For $X \sim \text{Pois}(\lambda)$, find $E(X!)$ (the average factorial of X), if it is finite.
2. Let $Z \sim \mathcal{N}(0,1)$. Find $E|Z|$.
3. Let $X \sim \text{Geom}(p)$ and let t be a constant. Find $E(e^{tX})$, as a function of t (this is known as the *moment generating function*; we will see later how this function is useful).

1) $X \sim Pois(\lambda)$

$E(X!)$ Let $g(x) = x!$

$$E(g(X)) = \sum_{x=0}^{k} g(x) f(x)$$ where $f(x)$ is PMF of X. and $0 \leq X \leq k$

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

$$E(g(x)) = \sum_{x=0}^{k} x! e^{-\lambda} \frac{\lambda^x}{x!}$$

$$= \sum_{x=0}^{k} e^{-\lambda} \lambda^x$$

$$= e^{-\lambda} \sum_{x=0}^{k} \lambda^x$$

$$= e^{-\lambda} \left(\frac{1-\lambda^{k+1}}{1-\lambda} \right) //$$

If $k \to \infty$, $E(g(x)) = e^{-\lambda} \left(\frac{1}{1-\lambda} \right)$

2) $Z \sim N(0,1)$.

$E(|Z|) = E(g(Z))$ where $g(x) = |x|$

$\int x f(x) \, dx$

$$\Rightarrow E(g(Z)) = \int_{-\infty}^{\infty} g(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$= \int_{0}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + \int_{0}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} x e^{-\frac{x^2}{2}} dx$$

Let $u = -\frac{x^2}{2}$ $\quad -\infty \leq u \leq 0$

$du = -x\,dx$

$dx = -\frac{1}{x} du$

$$-\frac{2}{\sqrt{2\pi}} \int_{0}^{-\infty} e^{u} du$$

$$= -\frac{2}{\sqrt{2\pi}} e^{u} \Big|_{0}^{-\infty} = \frac{2}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}} //$$

3) $X \sim Geom(p)$

$E(e^{tX})$ where t is constant

$$E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} (1-p)^x p$$

$$= p \sum_{x=0}^{\infty} [e^t (1-p)]^x$$

$$= p \left[\frac{1}{1 - e^t(1-p)} \right] = \frac{p}{1 - e^t(1-p)}$$

$$1 - e^t(1-p) = 0$$

$$e^t(1-p) = 1$$

$$-p = \frac{1}{e^t} - 1$$

$$p = 1 - \frac{1}{e^t}$$

I ♡ you

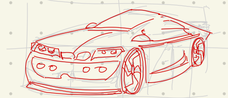
khag liang

1. A group of $n \geq 4$ people are comparing their birthdays (as usual, assume their birthdays are independent, are not February 29, etc.).

(a) Let I_{ij} be the indicator r.v. of i and j having the same birthday (for $i < j$). Is I_{12} independent of I_{34}? Is I_{12} independent of I_{13}? Are the I_{ij}'s independent?

(b) Explain why the Poisson Paradigm is applicable here even for moderate n, and use it to get a good approximation to the probability of at least 1 match when $n = 23$.

(c) About how many people are needed so that there is a 50% chance (or better) that two either have the same birthday or are only 1 day apart? (Note that this is much harder than the birthday problem to do exactly, but the Poisson Paradigm makes it possible to get fairly accurate approximations quickly.)



(a) I_{12} is independent of I_{34}. Knowing that 1, 2 have same birthdays tell nothing about the birthdays of 3, 4.

Same applies to. I_{12} and I_{13}

No $P(I_{13} \mid I_{12}=1, I_{23}=1) = 1$
$\neq P(I_{13})$
$\Rightarrow$ Not independent.

b) Number of indicator r.v is $\binom{n}{2}$ and each indicator r.v has small $p = \frac{1}{365}$ and not very dependent.

So Poisson works for this situation.

Let no. of matches be a r.v $X = \sum_{i=1}^{n} \sum_{j>i}^{n} I_{ij}$

$\lambda = \binom{n}{2} \frac{1}{365}$ For $n = 23$,

$$\lambda = \binom{23}{2} \cdot \frac{1}{365} = \frac{253}{365}$$

$P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$ where $\lambda = \frac{253}{365}$

$$P(X=0) = e^{-\lambda} \cdot 1 = e^{-253/365} = 0.4999 \approx 0.5$$

$$\Rightarrow P(X>0) = 1 - P(X=0) = 1/2 \;//$$

c) Let A_{ij} be an indicator r.v of i and j having same bday or being 1 day apart.

$$E(A_{ij}) = \frac{1}{365} + \frac{2}{365} = \frac{3}{365}$$

$$\lambda = \binom{n}{2} \frac{3}{365} = \frac{n!}{2!(n-2)!} \cdot \frac{3}{365}$$

Let Y be a r.v for the no. of pairs of ppl with same bday or 1 day apart.

$$P(Y=k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$P(Y=0) = e^{-\lambda} \leq \frac{1}{2}$$

$$e^{-\binom{n}{2}\frac{3}{365}} \leq \frac{1}{2}$$

$$-\binom{n}{2}\frac{3}{365} \leq \ln\left(\frac{1}{2}\right)$$

$$-\frac{n(n-1)}{2} \cdot \frac{3}{365} - \ln\frac{1}{2} \leq 0$$

$$\frac{n(n-1)}{2} \cdot \frac{3}{365} + \ln\frac{1}{2} \geq 0$$

$$\frac{3}{730}n^2 - \frac{3}{730}n + \ln\frac{1}{2} \geq 0$$

$n = 13.49$ or $n = -12.49$ (rej)

≈ 14

Check: $\binom{14}{2}\frac{3}{365} = \frac{273}{365}$

$$e^{-\frac{273}{365}} = 0.473 < 0.5 \checkmark.$$

$$e^{-\binom{13}{2}\frac{3}{365}} = 0.528 > 0.5 \checkmark$$

$\Rightarrow$ 14 is min no. of people for pair to occur.

2. Joe is waiting in continuous time for a book called *The Winds of Winter* to be released. Suppose that the waiting time T until news of the book's release is posted, measured in years relative to some starting point, has PDF $\frac{1}{5}e^{-t/5}$ for $t > 0$ (and 0 otherwise); this is known as the *Exponential distribution* with parameter 1/5. The news of the book's release will be posted on a certain website.

Joe is not so obsessive as to check multiple times a day; instead, he checks the website *once* at the end of each day. Therefore, he observes the day on which the news was posted, rather than the exact time T. Let X be this measurement, where $X = 0$ means that the news was posted within the first day (after the starting point), $X = 1$ means it was posted on the second day, etc. (assume that there are 365 days in a year). Find the PMF of X. Is this a distribution we have studied?

T : PDF $f(t) = \frac{1}{5}e^{-t/5}$

CDF $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} \frac{1}{5}e^{-t/5}\,dt$

$= -e^{-t/5}\Big|_0^x = -e^{-x/5} + e^0$

$= -e^{-x/5} + 1$

PMF of X : $P(X=k) = P(T \leq \frac{k+1}{365}) - P(T \leq \frac{k}{365})$

$= -e^{-\frac{k+1}{365}\cdot\frac{1}{5}} + 1 + e^{-\frac{k}{365}\cdot\frac{1}{5}} - 1$

$= e^{-\frac{k}{1825}} - e^{-\frac{k+1}{1825}}$

$= e^{-\frac{k}{1825}}\left[1 - e^{-\frac{1}{1825}}\right]$

$= \left(e^{-\frac{1}{1825}}\right)^k \left(1 - e^{-\frac{1}{1825}}\right)$

Let $p = 1 - e^{-\frac{1}{1825}}$

$P(X=k) = (1-p)^k p$

$\Rightarrow X \sim \text{Geom}(1 - e^{-\frac{1}{1825}})$ //

3. Let U be a Uniform r.v. on the interval $(-1, 1)$ (be careful about minus signs).

(a) Compute $E(U), \mathrm{Var}(U)$, and $E(U^4)$.

(b) Find the CDF and PDF of U^2. Is the distribution of U^2 Uniform on $(0, 1)$?

$E(U) = 0 \quad Var(U) = E(U^2) - E(U)^2$

$= E(U^2)$

$\frac{1}{2}\int_{-1}^{1} x^2 \, dx$

$= \frac{1}{2}\left[\frac{1}{3}x^3\right]_{-1}^{1} = \frac{1}{2}\left[\frac{1}{3} + \frac{1}{3}\right]$

$= \frac{1}{3}$ //

$E(U^4) = \frac{1}{2}\int x^4 \, dx$

$= \frac{1}{2}\left[\frac{1}{5}x^5\right]_{-1}^{1} = \frac{1}{2}\left[\frac{1}{5} - \frac{1}{5}\right]$

$= 0$ //

b) CDF of U^2: $P(U^2 \leq k)$

$= P(U \leq \pm\sqrt{k})$

$= P(-\sqrt{k} \leq U \leq \sqrt{k})$

$= \frac{2\sqrt{k}}{2} = \sqrt{k}$ //

PDF of U^2: $F(x) = \sqrt{x}$

$f(x) = \frac{1}{2}x^{-\frac{1}{2}} = \frac{1}{2\sqrt{x}}$

No. PDF is not constant.

4. Let F be a CDF which is continuous and strictly increasing. The inverse function, F^{-1}, is known as the *quantile function*, and has many applications in statistics and econometrics. Find the area under the curve of the quantile function from 0 to 1, in terms of the mean μ of the distribution F. Hint: Universality.

Let X be the r.v with CDF F.

$$F^{-1}(U) = X$$
$$U = F(X)$$

$$E(X) = E(F^{-1}(U))$$
$$= \int_0^1 F^{-1}(x) \cdot 1 \, dx$$
$$= \mu$$
$$\Rightarrow \int_0^1 F^{-1}(x)\,dx = \mu. \quad //$$

5)

5. Let $Z \sim \mathcal{N}(0,1)$. A measuring device is used to observe Z, but the device can only handle positive values, and gives a reading of 0 if $Z \leq 0$; this is an example of *censored data*. So assume that $X = ZI_{Z>0}$ is observed rather than Z, where $I_{Z>0}$ is the indicator of $Z > 0$. Find $E(X)$ and $\text{Var}(X)$.

$Z \sim N(0,1)$ with CDF $F(x)$ and PDF $f(x)$

$$X = ZI_{Z>0}$$

$$E(X) = \int_0^\infty x f(x)\,dx$$
$$= \int_0^\infty x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$$
$$= \frac{1}{\sqrt{2\pi}} \quad \text{from practice 4.2.}$$

$$\text{Var}(X) = E(X^2) - E(X)^2$$

$$E(X^2) = \frac{1}{2}\int_0^\infty x^2 f(x)\,dx$$
$$= \frac{1}{2}\int_0^\infty x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$$
$$= \frac{1}{2\sqrt{2\pi}} \int_0^\infty x^2 e^{-\frac{x^2}{2}}\,dx$$

Let $u = x$ $\qquad dv = xe^{-\frac{x^2}{2}}\,dx$

$du = dx$ $\qquad v = \int xe^{-\frac{x^2}{2}}\,dx$

$$= -e^{-\frac{x^2}{2}}$$

$$\Rightarrow E(X^2) = \frac{1}{2\sqrt{2\pi}}\left[-xe^{-\frac{x^2}{2}} + \int e^{-\frac{x^2}{2}}\,dx\right]$$
$$= \frac{1}{2\sqrt{2\pi}}\left[-xe^{-\frac{x^2}{2}}\Big|_0^\infty + \sqrt{2\pi}\right]$$
$$= \frac{1}{2}\left[-\lim_{x\to\infty} \frac{x}{e^{\frac{x^2}{2}}}\right] = \frac{1}{2}[0]$$
$$= \frac{1}{2}$$

$$\int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx = 1$$
$$\Rightarrow \int e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi}$$

$$\Rightarrow \quad Var(X) = \frac{1}{2} - \frac{1}{2\pi} = \frac{\pi - 1}{2\pi}$$