Do not submit

1. **Continuous versus discrete**

In class, a random variable X is specified by its *distribution* in the discrete case and by its *probability density function (pdf)* in the continuous case. To unify the two cases, we can define the *cumulative distribution function (cdf)* F, which is valid for both discrete and continuous random variables X, as follows:

$$F(a) = \Pr[X \leq a], \qquad a \in \mathbb{R}.$$

(a) In the discrete case, show that the cdf of a random variable contains exactly the same information as its distribution, by expressing F in terms of the distribution and expressing the distribution in terms of F. For simplicity, you may assume that the discrete random variable only takes on integer values.

Answer:

$$F(a) = \sum_{i=-\infty}^{a} \Pr[X = i]$$

$$\Pr[X = i] = F(i) - F(i-1)$$

(b) In the continuous case, show that the cdf of a random variable contains exactly the same information as its pdf, by expressing F in terms of the pdf and expressing the pdf in terms of F.

Answer:

$$F(x) = \int_{-\infty}^{x} f(a)da$$

$$f(x) = \frac{d}{dx}F(x)$$

(c) Identify two key properties that a cdf of any random variable has to satisfy.

Answer:

i. $0 \leq F(a) \leq 1$
ii. $\lim_{a\to-\infty} F(a) = 0$
iii. $\lim_{a\to\infty} F(a) = 1$

iv. $\forall \varepsilon > 0,\ F(a) \leq F(a+\varepsilon)$

In other words, a cdf of any random variable must rise monotonically from 0 to 1 (this is because probabilities range from 0 to 1)

2. Discrete and continuous random variables have a lot of similarities but some differences too.

   (a) Suppose X is a discrete random variable. Let $Y = cX$ for some constant c. Express the distribution of Y in terms of the distribution of X.

   Answer:

   $$\forall i \in \mathbb{Z},\ \Pr[Y = ci] = \Pr[X = i]$$

   (b) Suppose X is a continuous random variable. Let $Y = cX$ for some constant c. Express the pdf of Y in terms of the pdf of X. Is there any difference with the discrete case? (Hint: work with cdf's.)

   Answer: Let Y have cdf F_Y and pdf f_Y, and X have cdf F_X and pdf f_X. Since $Y = cX$, we can see that $\forall i \in \mathbb{R},\ F_Y(ic) = F_X(i)$. From our answer in part (7b), we can then find f_Y.

   $$\begin{aligned} f_Y(x) &= \frac{d}{dx}F_Y(x) \\ &= \frac{d}{dx}F_X(x/c) \\ &= \frac{1}{c}\cdot f_X(x/c) \qquad \text{(By the chain rule from calculus)} \end{aligned}$$

   We can see quite clearly that our continuous f_Y has an additional factor of $\frac{1}{c}$ that our discrete Y does not.

   (c) If $X = N(\mu,\sigma^2)$, what is the density of $Y = cX$?

   Answer: Let Y have pdf f_Y and X have pdf $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}$. Since $f_Y(x) = \frac{1}{c}\cdot f_X(x/c)$, we have $Y \sim N(c\mu,(c\sigma)^2)$, so $f_Y(x) = \dfrac{1}{c\sqrt{2\pi\sigma^2}}e^{-(x/c-\mu)^2/2\sigma^2}$.

3. We begin by proving two very useful properties of the exponential distribution. We then use them to solve a problem in digital photography.

   (a) Let random variable X have exponential distribution with parameter λ. Show that, for any positive s,t, we have

   $$\Pr\left[X > s+t \,\middle|\, X > t\right] = \Pr[X > s].$$

   This is the "memoryless" property of the exponential distribution. We already saw the analogous memoryless property of the geometric distribution in the section 11 worksheet.

   Answer: If X has an exponential distribution with parameter λ, then we know that $\Pr[X > a] = \int_a^\infty \lambda e^{-\lambda x}dx = e^{-\lambda a}$, so

   $$\begin{aligned} \Pr\left[X > s+t \,\middle|\, X > t\right] &= \Pr\left[X > s+t \cap X > t\right]/\Pr[X > t] \\ &= \Pr[X > s+t]/\Pr[X > t] \\ &= e^{-\lambda(s+t)}/e^{-\lambda t} \\ &= e^{-\lambda s} \\ &= \Pr[X > s] \end{aligned}$$

(b) Let random variables X_1, X_2 be independent and exponentially distributed with parameters λ_1, λ_2. Show that the random variable $Y = \min\{X_1, X_2\}$ is exponentially distributed with parameter $\lambda_1 + \lambda_2$. (Hint: work with cdf's.)

Answer: Using the hint, we'll work with the cdf's of these random variables.

$$
\begin{aligned}
\Pr[Y \leq a] &= 1 - \Pr[Y > a] \\
&= 1 - \Pr[\min\{X_1, X_2\} > a] \\
&= 1 - \Pr[X_1 > a \cap X_2 > a] \\
&= 1 - \Pr[X_1 > a] \Pr[X_2 > a] \\
&= 1 - (e^{-\lambda_1 a})(e^{-\lambda_2 a}) \\
&= 1 - e^{-(\lambda_1 + \lambda_2)a}
\end{aligned}
$$

This shows that Y has an exponential distribution with parameter $\lambda_1 + \lambda_2$.

(c) You have a digital camera that requires two batteries to operate. You purchase n batteries, labeled $1, 2, \ldots, n$, each of which has a lifetime that is exponentially distributed with parameter λ and is independent of all the 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.

Answer: Let T be the time when we are left with only one battery. We want to compute $E[T]$.To facilitate the calculation, we can break up T into the intervals

$$T = T_1 + T_2 + \ldots + T_{n-1}$$

where T_i is the time between the $(i-1)^{st}$ failure and i^{th} failure. After $n-1$ failures, we are left with only one battery and the process ends. Note that the i^{th} failure *need not* correspond to the battery labeled i.

We can calculate $E[T]$ as

$$E[T] = E[T_1] + E[T_2] + \ldots + E[T_{n-1}]$$

It now remains to find $E[T_i]$ for each i. To calculate $E[T_i]$, note that after $i-1$ failures, we have the battery labeled i and another battery labeled j for $j < i$ installed in the camera.

The lifetime of the battery labeled i is distributed as an exponential random variable with parameter λ. Also, because of the memoryless property proved in part (a), the *remaining* lifetime of the battery labeled j, conditioned on the battery being functional till the time of the $(i-1)^{st}$ failure, is *also* distributed as an exponential random variable with parameter λ. Hence, T_i is the minimum of two exponential random variables, each with parameter λ. By part (b), T_i is exponentially distributed with parameter 2λ. This gives

$$E[T_i] = \frac{1}{2\lambda}$$

and hence,

$$E[T] = E[T_1] + E[T_2] + \ldots + E[T_{n-1}] = \frac{n-1}{2\lambda}$$

(d) In the scenario of part c, what is the probability that battery i is the last remaining working battery, as a function of i?

Answer: We first want to calculate, given that the battery labeled i and the battery labeled j, (for some $j \neq i$) are installed, the probability that the battery labeled j fails first. If neither of the batteries have failed, then conditioned on this fact, the distributions of their remaining lifetimes are identical (exponential with parameter λ). Hence, by symmetry, the probability that the one labeled j fails first is 1/2.

For batteries $2,3,\ldots,n$, the battery labelled i is installed after $i-2$ batteries have failed. The probability that it is the last one left, is the probability that in each of the remaining $n-i+1$ failures that remain, it is the other installed battery in the camera which fails first and not the battery labeled i. From the previous calculation, each of these events happen with probability $\frac{1}{2}$, conditioned on the history. Hence, the probability that the battery labelled i is the last one left is $1/2^{n-i+1}$. However, this formula is slightly off for the battery labeled 1; like battery 2, it also needs to survive $n-1$ failures, not n failures, and so we get:

$$\Pr[\text{last battery is battery } i] = \begin{cases} \frac{1}{2^{n-1}} & i = 1 \\ \frac{1}{2^{n-i+1}} & 2 \leq i \leq n \\ 0 & \text{otherwise} \end{cases}$$

4. Suppose a set of final grades for a course are approximately normally distributed with a mean of 64 and a standard deviation of 7.1. (You are free to use the facts that $\Pr[N(0,1) \leq 1.3] \approx 0.9$ and $\Pr[N(0,1) \leq 1.65] \approx 0.95$.)

(a) Find the lowest passing grade if the bottom 5% of the students fail the class.

Answer: Since $\Pr[N(0,1) \leq 1.65] \approx 0.95$, $\Pr[N(\mu,\sigma^2) \leq \mu - 1.65\sigma] \approx 0.05$. Our distribution has $\mu = 64$ and $\sigma = 7.1$, so $64 - 1.65 \cdot 7.1 \approx 52.3$ is the lowest passing grade.

(b) Find the highest B+ if the top 10% of the students are given A's or A-'s.

Answer: Since $\Pr[N(0,1) \leq 1.3] \approx 0.9$, $\Pr[N(\mu,\sigma^2) \leq \mu + 1.3\sigma] \approx 0.90$. Again, our distribution has $\mu = 64$ and $\sigma = 7.1$, so $64 + 1.3 \cdot 7.1 \approx 73.2$ is the B+ to A- borderline.

5. **Central Limit Theorem**

Suppose you roll a standard die 2000 times and let X be the sum of the values you get. Using the Central Limit Theorem, for what value of a is $\Pr[X \geq a] \approx \Pr[N(0,1) \geq 2]$? Justify your answer.

Answer: The Central Limit Theorem lets us estimate $\Pr[X \geq a] \approx \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^\infty e^{-(x-\mu)^2/2\sigma^2} dx = \Pr[N(\mu,\sigma^2) \geq a] = \Pr[N(0,1) \geq \frac{a-\mu}{\sigma}]$, where μ is the mean (expectation) of X and σ is the standard deviation of X. Thus, we must find the expectation and standard deviation of X. Since $X = \sum_{i=1}^{2000} X_i$, where X_i is the value of the i^{th} roll, $\mathrm{E}(X) = \sum_{i=1}^{2000} \mathrm{E}(X_i) = 2000 \cdot 3.5 = 7000$. Now we need to find the variance of X to get the stardard deviation. Since the die rolls are independent, we can use linearity of variance to find that $\mathrm{Var}(X) = \sum_{i=1}^{2000} \mathrm{Var}(X_i) = 2000 \cdot \frac{35}{12} \approx 5833$. Thus, our standard deviation is $\sigma = \sqrt{5833} \approx 76.38$. Now, since $\frac{a-\mu}{\sigma} = 2$, we solve for a and find $a = 2\sigma + \mu \approx 2 \cdot 76.38 + 7000 = 7152.76$.