ECE 302 Spring 2012

Practice problems: Continuous random variables; uniform, exponential, normal, lognormal, Rayleigh, Cauchy, Pareto, Gaussian mixture, Erlang, and Laplace random variables; quantization; optimal portfolios; simulating a fair coin with a biased one; dependence and correlation; binary detection.

Ilya Pollak

These problems have been constructed over many years, using many different sources. If you find that some problem or solution is incorrectly attributed, please let me know at ipollak@ecn.purdue.edu.

Suggested reading: Sections 3.1-3.7, 4.2, and 4.3 in the recommended text [1]. Equivalently, Sections 3.1-3.6, 4.1-4.5, 4.7 (correlation and covariance only) in the Leon-Garcia text [3].

Problem 1. X is a continuous random variable, uniformly distributed between 10 and 12. Find the CDF, the PDF, the mean, the variance, and the standard deviation of the random variable $Y = X^2$.

Solution. Since X is between 10 and 12 with probability 1, it follows that $Y = X^2$ is between 100 and 144 with probability 1. Hence, the CDF of Y is zero for all outcomes below 100 and one for all outcomes above 144. Between 100 and 144, we have:

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(X^2 \le y) = \mathbf{P}(X \le \sqrt{y}) = F_X(\sqrt{y}) = \frac{\sqrt{y}-10}{2},$$

where we used the fact derived in class that for $x \in [10, 12]$,

$$F_X(x) = \frac{x-10}{12-10}.$$

The answer for the CDF of Y:

$$F_Y(y) = \begin{cases} 0, & y \le 100 \\ \frac{\sqrt{y}-10}{2}, & 100 \le y \le 144 \\ 1, & y \ge 144. \end{cases}$$

The PDF of Y is obtained by differentiating the CDF:

$$f_Y(y) = \begin{cases} 0, & y \le 10 \\ \frac{1}{4\sqrt{y}} & 100 \le y \le 144 \\ 0, & y \ge 144. \end{cases}$$

The mean and the variance of Y can be computed either using the PDF of X or using the newly computed PDF of Y. As derived in class, the mean and variance of X are 11 and 1/3, respectively. This means that $E[Y] = E[X^2] = \text{var}(X) + (E[X])^2 = 121\frac{1}{3}$. Another way of obtaining the same result is through defining a new random variable

$$S = \frac{X-11}{2}.$$

Then S is uniformly distributed between -1/2 and 1/2, and has mean zero, as shown in class. It was also shown in class that $E[S^2] = 1/12$. Since $X = 2S + 11$, we have:

$$Y = X^2 = 4S^2 + 44S + 121,$$

and

$$E[Y] = E[X^2] = 4E[S^2] + 44E[S] + 121 = 121\frac{1}{3}.$$

We can also obtain the same result by using the PDF of Y:

$$\begin{aligned}
E[Y] &= \int_{100}^{144} y\frac{1}{4\sqrt{y}}dy \\
&= \int_{100}^{144} \frac{\sqrt{y}}{4}dy \\
&= \left.\frac{y^{3/2}}{6}\right|_{100}^{144} \\
&= \frac{1728 - 1000}{6} = \frac{364}{3} = 121\frac{1}{3}.
\end{aligned}$$

To compute the variance of Y, we can use the formula $\mathrm{var}(Y) = E[Y^2] - (E[Y])^2$. In order to use this formula, we need to compute $E[Y^2]$ first. We can do this using the PDF of Y:

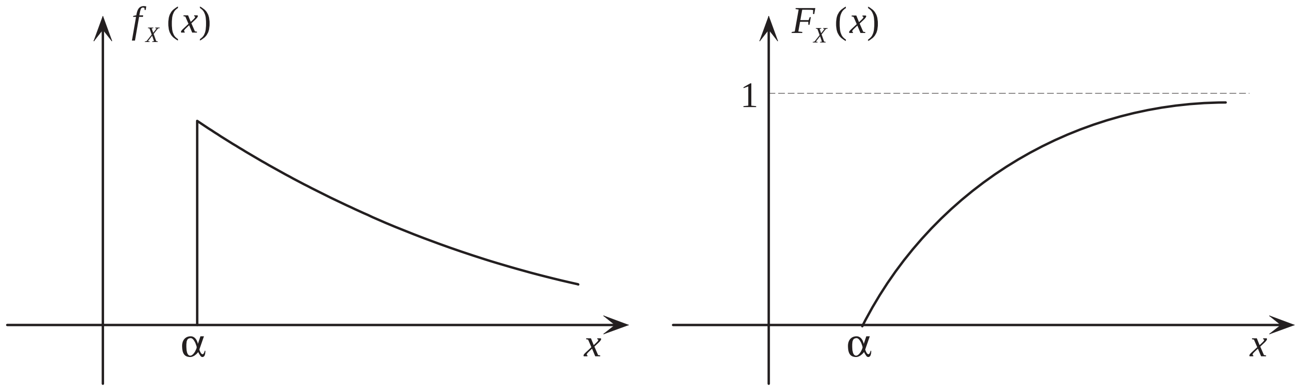
$$\begin{aligned}
E[Y^2] &= \int_{100}^{144} y^2\frac{1}{4\sqrt{y}}dy \\
&= \int_{100}^{144} \frac{y^{3/2}}{4}dy \\
&= \left.\frac{y^{5/2}}{10}\right|_{100}^{144} \\
&= \frac{12^5 - 10^5}{10} = \frac{248832 - 100000}{10} = \frac{74416}{5} = 14883.2
\end{aligned}$$

Alternatively, we can use the PDF of X to obtain the same result:

$$\begin{aligned}
E[Y^2] = E[X^4] &= \int_{10}^{12} \frac{x^4}{2}dx \\
&= \left.\frac{x^5}{10}\right|_{10}^{12} \\
&= \frac{12^5 - 10^5}{10} = 14883.2
\end{aligned}$$

Now we can compute the variance of Y:

$$\begin{aligned}
\mathrm{var}(Y) &= E[Y^2] - (E[Y])^2 = \frac{74416}{5} - \left(\frac{364}{3}\right)^2 = \frac{74416 \cdot 9 - 364^2 \cdot 5}{45} = \frac{669744 - 662480}{45} \\
&= \frac{7264}{45} \approx 161.4
\end{aligned}$$



(a) PDF of X (b) CDF of X

Figure 1: PDF and CDF of X

The standard deviation is the square root of the variance:

$$\sigma_Y = \sqrt{\frac{7264}{45}} \approx \sqrt{161.4} \approx 12.7.$$

Problem 2. *(Ilya Pollak and Bin Ni.)*
A random variable X is called a *shifted exponential* when its PDF has the following form:

$$f_X(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x-\alpha}{\theta}}, & x > \alpha \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the CDF of X.

Solution. Define $Y = X - \alpha$. The cumulative distribution function of this random variable is then:

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(X - \alpha \le y) = \mathbf{P}(X \le y + \alpha) = F_X(y + \alpha). \tag{1}$$

Differentiating the leftmost part of this equation and the rightmost part, we get $f_Y(y)$ and $f_X(y + \alpha)$, respectively:

$$f_Y(y) = f_X(y + \alpha) = \begin{cases} \frac{1}{\theta} e^{-\frac{y}{\theta}}, & y > 0 \\ 0, & \text{otherwise,} \end{cases}$$

i.e. Y is an exponential random variable with mean θ. Using Eq. (1), as well as the exponential CDF derived in class, we have:

$$F_X(x) = F_Y(x - \alpha) = \begin{cases} 1 - e^{-\frac{x-\alpha}{\theta}}, & x > \alpha \\ 0, & \text{otherwise.} \end{cases}$$

$F_X(x)$ is depicted in Figure 1(b).

(b) Calculate the mean and the variance of X.

Solution. It was derived in class that the mean and variance of the exponential random variable Y we defined in the last part are θ and θ^2 respectively. Since α is just a constant, we have:

$$\begin{aligned} E[X] &= E[Y+\alpha] = E[Y] + \alpha = \theta + \alpha \\ Var[X] &= Var[Y+\alpha] = Var[Y] = \theta^2 \end{aligned}$$

(c) Find the real number μ that satisfies: $F_X(\mu) = 1/2$. This number μ is called the *median* of the random variable X.

Solution. With the CDF we got in Part(a), we have:

$$\begin{aligned} 1 - e^{-\frac{\mu-\alpha}{\theta}} &= \frac{1}{2} \\ \Rightarrow \frac{\mu-\alpha}{\theta} &= \ln 2 \\ \Rightarrow \mu &= \theta \ln 2 + \alpha. \end{aligned}$$

Problem 3. LOGNORMAL RANDOM VARIABLE, ITS PDF, MEAN, AND MEDIAN. *(Ilya Pollak).* Suppose Y is a normal random variable with mean μ and standard deviation σ. Let $X = e^Y$. X is called a *lognormal* random variable since its log is a normal random variable.

(a) Find the PDF and the expectation of X.

(b) The *median* of a continuous random variable Z is defined as such number d that $\mathbf{P}(Z \le d) = \mathbf{P}(Z \ge d) = 1/2$. Find the median of X.

(c) Lognormal random variables are widely used in finance and economics to model compounded returns of various assets. Suppose S_k and S_n are the prices of a financial instrument on days k and n, respectively. For $k < n$, *the gross return* $G_{k,n}$ between days k and n is defined as $G_{k,n} = S_n/S_k$ and is equal to the amount of money you would have on day n if you invested \$1 on day k. For example, $G_{k,n} = 1.05$ means that an investment of \$1 on day k would earn five cents by day n.

(i) Show that $G_{1,n} = G_{1,2} \cdot G_{2,3} \cdot \ldots \cdot G_{n-1,n}$. In other words, the total gross return from day 1 to day n is equal to the product of all daily gross returns over this time period.

(ii) Show that if $G_{1,2},\ G_{2,3}, \ldots, G_{n-1,n}$ are independent lognormal random variables, then $G_{1,n}$ is lognormal. You can use the fact that the sum of independent normal random variables is normal.

(iii) Suppose that $G_{1,2},\ G_{2,3}, \ldots, G_{n-1,n}$ are independent lognormal random variables, each with mean g. Find the mean of $G_{1,n}$.

(iv) Suppose that $G_{1,2},\ G_{2,3}, \ldots, G_{n-1,n}$ are independent lognormal random variables, each with mean g and variance v. Find the variance of $G_{1,n}$.

Solution.

(a) From the definition of X it follows that $\mathbf{P}(X \le 0) = 0$. For $x > 0$, we have:

$$F_X(x) = \mathbf{P}(X \le x) = \mathbf{P}\left(e^Y \le x\right) = \mathbf{P}(Y \le \log x) = F_Y(\log x) \tag{2}$$

Therefore, for $x > 0$,

$$\begin{aligned} f_X(x) &= F_X'(x) = \frac{d}{dx}F_Y(\log x) = \frac{1}{x}f_Y(\log x) \\ &= \frac{1}{x\sqrt{2\pi}\sigma}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} \end{aligned}$$

(b) The mean of X can be found as follows:

$$\begin{aligned} E[X] &= E\left[e^Y\right] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}e^y e^{-\frac{(y-\mu)^2}{2\sigma^2}}\,dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y^2-2y\mu+\mu^2-2y\sigma^2}{2\sigma^2}\right)dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y^2-2y(\mu+\sigma^2)+\mu^2+2\mu\sigma^2+\sigma^4-2\mu\sigma^2-\sigma^4}{2\sigma^2}\right)dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu-\sigma^2)^2-2\mu\sigma^2-\sigma^4}{2\sigma^2}\right)dy \\ &= \exp\left(\mu+\frac{\sigma^2}{2}\right)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu-\sigma^2)^2}{2\sigma^2}\right)dy \\ &= \exp\left(\mu+\frac{\sigma^2}{2}\right), \end{aligned}$$

where we used the fact that the last integral is the integral of a normal PDF with mean $\mu+\sigma^2$ and standard deviation σ, and hence is equal to one.

The median d of X satisfies $F_X(d) = 1/2$. But, on the other hand, it was shown in Eq. (2) that $F_X(d) = F_Y(\log d)$. Since a normal PDF is symmetric about its mean, we know that $F_Y(\mu) = 1/2$. Hence, $\log d = \mu$, and

$$d = e^\mu$$

Note that the lognormal distribution is skewed to the right (see Fig. 2) and hence its mean is larger than the median. Fig. 2 is a plot of the lognormal PDF with $\mu = \log(20000) \approx 9.9$ and $\sigma = 1$. The mean and the median for this PDF are $20000\sqrt{e} \approx 32974$ and 20000, respectively.

(c) Using the definition of gross returns in terms of prices, we have:

$$G_{1,2}\cdot G_{2,3}\cdot \ldots \cdot G_{n-1,n} = \frac{S_2}{S_1}\cdot\frac{S_3}{S_2}\cdot\ldots\cdot\frac{S_n}{S_{n-1}} = \frac{S_n}{S_1} = G_{1,n}.$$

Taking logs, we have:

$$\log G_{1,n} = \log G_{1,2} + \log G_{2,3} + \ldots + \log G_{n-1,n}.$$

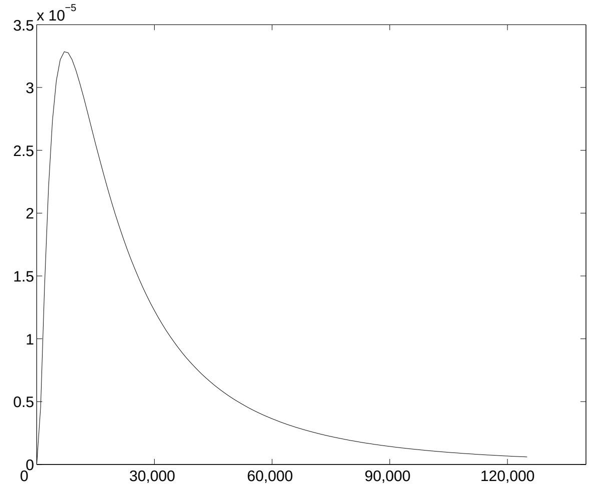


Figure 2: Lognormal distribution with parameters $\mu = \log(20000) \approx 9.9$ and $\sigma = 1$.

If $G_{1,2},\ G_{2,3}, \ldots, \log G_{n-1,n}$ are independent and lognormal, then $\log G_{1,2},\ \log G_{2,3}, \ldots, \log G_{n-1,n}$ are independent and normal, and therefore their sum $\log G_{1,n}$ is normal, and so $G_{1,n}$ is lognormal. Using the independence of the gross returns, we have:

$$\begin{aligned} E\left[G_{1,n}\right] &= E\left[G_{1,2} \cdot G_{2,3} \cdot \ldots \cdot G_{n-1,n}\right] \\ &= E\left[G_{1,2}\right] \cdot E\left[G_{2,3} \cdot \ldots \cdot\right] E\left[G_{n-1,n}\right] \\ &= g^{n-1}. \end{aligned}$$

$$\begin{aligned} E\left[G_{1,n}^2\right] &= E\left[G_{1,2}^2 \cdot G_{2,3}^2 \cdot \ldots \cdot G_{n-1,n}^2\right] \\ &= E\left[G_{1,2}^2\right] \cdot E\left[G_{2,3}^2 \cdot \ldots \cdot\right] E\left[G_{n-1,n}^2\right] \\ &= (v+g^2)^{n-1}. \end{aligned}$$

The last equality came from the fact that, for any random variable V, $\mathrm{var}(V) = E[V^2] - (E[V])^2$, and therefore, using $V = G_{k,k+1}^2$, we obtain $E\left[G_{k,k+1}^2\right] = v + g^2$. Using this identity again for $V = G_{1,n}$, we have:

$$\begin{aligned} \mathrm{var}\left(G_{1,n}\right) &= E\left[G_{1,n}^2\right] - \left(E\left[G_{1,n}\right]\right)^2 \\ &= (v+g^2)^{n-1} - g^{2(n-1)}. \end{aligned}$$

Problem 4. On exam taking strategies. *(Ilya Pollak.)*
This problem illustrates some probabilistic tools which may be used in analyzing and comparing various decision strategies.

Suppose you are taking a course, Decisions 302, taught by Prof. Kind. The course has three tests. Apart from some exceptions described below, your final score S is computed as

$$S = \frac{1}{3}(T_1 + T_2 + T_3),$$

where T_n is your score on test n, with $n = 1, 2, 3$. The final grade in the course is then computed as follows:

$$\begin{array}{rll} \text{You get an A} & \text{if} & S \geq 90, \\ \text{B} & \text{if} & 80 \leq S < 90 \\ \text{C} & \text{if} & 70 \leq S < 80 \\ \text{D} & \text{if} & 60 \leq S < 70 \\ \text{F} & \text{if} & S < 60 \end{array}$$

Your intrinsic knowledge of the subject is x. I.e., if the tests were perfectly designed and your scores always perfectly reflected your knowledge, you would get a score of x on each of the three tests. However, due to various random factors (such as imperfect design of the tests, possibility of random errors or lucky guesses on your part, etc.), your actual scores T_n are normal random variables with mean x and variance σ^2. (We are assuming here that test scores are real numbers between $-\infty$ and ∞.) It is reasonable to suppose that the random factors influencing the three test results are independent, and therefore we assume that the random variables T_1, T_2, and T_3 are independent.

Before Exam 1, Prof. Kind allows you to decide to skip Exam 1. If you decide to skip it, your final score will be S', computed as follows:

$$S' = \frac{1}{2}(T_2 + T_3).$$

Note that this decision must be made *before* you know any of your test scores. If you decide to take Exam 1, your final score will be S. If you decide to skip Exam 1, your final score will be S'.

(a) Suppose that your intrinsic knowledge is $x = 91$ and standard deviation is $\sigma = 3$. Which of the two possibilities should you choose in order to maximize your probability of getting an A in the course? Here and elsewhere, you can use the fact that the sum of several independent normal random variables is a normal random variable.

(b) Suppose that your intrinsic knowledge is $x = 55$ and standard deviation is $\sigma = 5$. Which of the two possibilities should you choose in order to minimize your probability of getting an F in the course?

(c) Suppose that your intrinsic knowledge is $x = 85$ and standard deviation is $\sigma = 5$. Which of the two possibilities should you choose in order to maximize your probability of getting a B in the course?

(d) Suppose again that your intrinsic knowledge is $x = 85$ and standard deviation is $\sigma = 5$. Which of the two possibilities should you choose in order to maximize your probability of getting an A in the course? Are there any risks associated with this choice?

Solution. Both S and S' are normal since each of them is a linear combination of independent normal random variables. We denote their means by μ_S and $\mu_{S'}$ and their standard deviations by σ_S and $\sigma_{S'}$, respectively. Using the linearity of expectations, the means are:

$$\begin{aligned}
\mu_S &= \frac{1}{3}(E[T_1] + E[T_2] + E[T_3]) = \frac{1}{3}(x + x + x) = x, \\
\mu_{S'} &= \frac{1}{2}(E[T_2] + E[T_3]) = \frac{1}{2}(x + x) = x.
\end{aligned}$$

Since the random variables T_1, T_2, and T_3 are independent, the variance of their sum is the sum of their variances. In addition, we use the fact that the variance of aY is $a^2\mathrm{var}(Y)$ for any number a and any random variable Y:

$$\begin{aligned}
\sigma_S &= \sqrt{\left(\frac{1}{3}\right)^2 (\mathrm{var}(T_1) + \mathrm{var}(T_2) + \mathrm{var}(T_3))} = \sqrt{\left(\frac{1}{3}\right)^2 (\sigma^2 + \sigma^2 + \sigma^2)} = \frac{\sigma}{\sqrt{3}} \\
\sigma_{S'} &= \sqrt{\left(\frac{1}{2}\right)^2 (\mathrm{var}(T_2) + \mathrm{var}(T_3))} = \sqrt{\left(\frac{1}{2}\right)^2 (\sigma^2 + \sigma^2)} = \frac{\sigma}{\sqrt{2}}
\end{aligned}$$

Thus, the two random variables have identical means equal to your intrinsic knowledge x, but S has a smaller variance than S'. Thus, if your objective is to increase your probability of getting the final score which is very close to your intrinsic knowledge, you should choose Option 1. If your objective is to have as large a chance as possible to get a score which is far from your intrinsic knowledge, you should choose Option 2. In other words, if you are a very good student, the best thing to do is to take as many exams as possible. This will reduce the measurement noise and increase the chances that the grade you get is the grade that you deserve. If you are a bad student, the best thing to do is to take as few exams as possible. This will *increase* the measurement noise and increase the chances that you do *not* get the grade that you deserve but get a higher grade. If you are an average student, your best strategy depends on your appetite for risk. If you like taking on more risk of a lower-than-deserved grade in order to give yourself a larger probability of a higher-than-deserved grade, you should choose Option 2. If you are risk-averse and just want to maximize the probability of the grade you deserve, you should choose Option 1.

This qualitative discussion is reflected in the following solutions to the four parts of the problem.

(a) The probabilities of getting an A under the two strategies are $\mathbf{P}(S \geq 90)$ and $\mathbf{P}(S' \geq 90)$:

$$\begin{aligned}
\mathbf{P}(S \geq 90) &= \mathbf{P}\left(\frac{S - \mu_S}{\sigma_S} \geq \frac{90 - \mu_S}{\sigma_S}\right) \\
&= \mathbf{P}\left(\frac{S - 91}{\sqrt{3}} \geq \frac{90 - 91}{\sqrt{3}}\right) \\
&= \mathbf{P}\left(\frac{S - 91}{\sqrt{3}} \geq -\frac{1}{\sqrt{3}}\right) \\
&= 1 - \Phi\left(-\frac{1}{\sqrt{3}}\right) \\
&\approx 1 - \Phi(-0.58) \\
&\approx 0.72
\end{aligned}$$

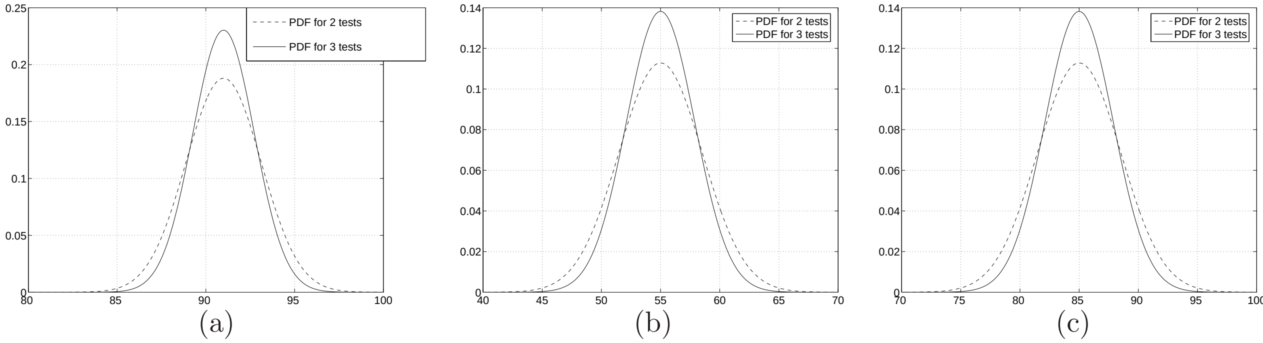


Figure 3: Exam taking problem: panel (a) shows the PDFs for Part (a), panel (b) shows the PDFs for Part (b), and panel (c) shows the PDFs for Parts (c) and (d).

$$\begin{aligned}
\mathbf{P}(S' \geq 90) &= \mathbf{P}\left(\frac{S'-\mu_{S'}}{\sigma_{S'}} \geq \frac{90-\mu_{S'}}{\sigma_{S'}}\right) \\
&= \mathbf{P}\left(\frac{S'-91}{3/\sqrt{2}} \geq \frac{90-91}{3/\sqrt{2}}\right) \\
&= \mathbf{P}\left(\frac{S'-91}{3/\sqrt{2}} \geq -\frac{\sqrt{2}}{3}\right) \\
&= 1-\Phi\left(-\frac{\sqrt{2}}{3}\right) \\
&\approx 1-\Phi(-0.47) \\
&\approx 0.68 < 0.72
\end{aligned}$$

Thus, Option 1 gives a higher probability of an A than Option 2. In these calculations, we have used the fact that both $\frac{S-\mu_S}{\sigma_S}$ and $\frac{S'-\mu_{S'}}{\sigma_{S'}}$ are standard normal random variables, i.e., have zero mean and unit variance.

This is illustrated in Fig. 3(a). The probabilities to get an A are the areas under the two PDF curves to the right of 90. The solid curve, which is the PDF for the final score under the three-test strategy, has a larger probability mass to the right of 90 than the dashed curve, which is the PDF for the final score under the two-test strategy.

(b) The probabilities of getting an F under the two strategies are $\mathbf{P}(S < 60)$ and $\mathbf{P}(S' < 60)$:

$$\begin{aligned}
\mathbf{P}(S < 60) &= \mathbf{P}\left(\frac{S-\mu_S}{\sigma_S} < \frac{60-\mu_S}{\sigma_S}\right) \\
&= \mathbf{P}\left(\frac{S-55}{5/\sqrt{3}} < \frac{60-55}{5/\sqrt{3}}\right) \\
&= \mathbf{P}\left(\frac{S-55}{5/\sqrt{3}} < \sqrt{3}\right) \\
&= \Phi(\sqrt{3})
\end{aligned}$$

$$
\begin{aligned}
&\approx \Phi(1.73) \\
&\approx 0.96 \\
\mathbf{P}(S' < 60) &= \mathbf{P}\left(\frac{S' - \mu_{S'}}{\sigma_{S'}} < \frac{60 - \mu_{S'}}{\sigma_{S'}}\right) \\
&= \mathbf{P}\left(\frac{S' - 55}{5/\sqrt{2}} < \frac{60 - 55}{5/\sqrt{2}}\right) \\
&= \mathbf{P}\left(\frac{S' - 55}{5/\sqrt{2}} < \sqrt{2}\right) \\
&= \Phi(\sqrt{2}) \\
&\approx \Phi(1.41) \\
&\approx 0.92 < 0.96
\end{aligned}
$$

Thus, Option 2 gives a lower probability of an F than Option 1. In these calculations, we again have used the fact that both $\frac{S-\mu_S}{\sigma_S}$ and $\frac{S'-\mu_{S'}}{\sigma_{S'}}$ are standard normal random variables, i.e., have zero mean and unit variance.

This is illustrated in Fig. 3(b). The probabilities to get an F are the areas under the two PDF curves to the left of 60. The solid curve, which is the PDF for the final score under the three-test strategy, has a larger probability mass to the left of 60 than the dashed curve, which is the PDF for the final score under the two-test strategy.

(c) The probabilities of getting a B under the two strategies are $\mathbf{P}(80 \le S < 90)$ and $\mathbf{P}(80 \le S' < 90)$:

$$
\begin{aligned}
\mathbf{P}(80 \le S < 90) &= \mathbf{P}\left(\frac{80 - \mu_S}{\sigma_S} \le \frac{S - \mu_S}{\sigma_S} < \frac{90 - \mu_S}{\sigma_S}\right) \\
&= \mathbf{P}\left(\frac{80 - 85}{5/\sqrt{3}} \le \frac{S - 85}{5/\sqrt{3}} < \frac{90 - 85}{5/\sqrt{3}}\right) \\
&= \mathbf{P}\left(-\sqrt{3} \le \frac{S - 85}{5/\sqrt{3}} < \sqrt{3}\right) \\
&= \Phi(\sqrt{3}) - \Phi(-\sqrt{3}) \\
&= \Phi(\sqrt{3}) - (1 - \Phi(\sqrt{3})) \\
&= 2\Phi(\sqrt{3}) - 1 \\
&\approx 2\Phi(1.73) - 1 \\
&\approx 2 \cdot 0.96 - 1 = 0.92 \\
\mathbf{P}(80 \le S' < 90) &= \mathbf{P}\left(\frac{80 - \mu_{S'}}{\sigma_{S'}} \le \frac{S' - \mu_{S'}}{\sigma_{S'}} < \frac{90 - \mu_{S'}}{\sigma_{S'}}\right) \\
&= \mathbf{P}\left(\frac{80 - 85}{5/\sqrt{2}} \le \frac{S' - 85}{5/\sqrt{2}} < \frac{90 - 85}{5/\sqrt{2}}\right) \\
&= \mathbf{P}\left(-\sqrt{2} \le \frac{S' - 85}{5/\sqrt{2}} < \sqrt{2}\right) \\
&= \Phi(\sqrt{2}) - \Phi(-\sqrt{2})
\end{aligned}
$$

$$\begin{aligned}
&= \Phi(\sqrt{2}) - (1 - \Phi(\sqrt{2})) \\
&= 2\Phi(\sqrt{2}) - 1 \\
&\approx 2\Phi(1.41) - 1 \\
&\approx 2 \cdot 0.92 - 1 = 0.84 < 0.92
\end{aligned}$$

Thus, Option 1 gives a higher probability of an B than Option 2. This is illustrated in Fig. 3(c). The probabilities to get a B are the areas under the two PDF curves, between 80 and 90. The solid curve, which is the PDF for the final score under the three-test strategy, has a larger probability mass between 80 and 90 than the dashed curve, which is the PDF for the final score under the two-test strategy.

(d) As stated in Part (a), the probabilities of getting an A under the two strategies are $\mathbf{P}(S \geq 90)$ and $\mathbf{P}(S' \geq 90)$. We recompute these for the new mean and variance.

$$\begin{aligned}
\mathbf{P}(S \geq 90) &= \mathbf{P}\left(\frac{S - \mu_S}{\sigma_S} \geq \frac{90 - \mu_S}{\sigma_S}\right) \\
&= \mathbf{P}\left(\frac{S - 85}{5/\sqrt{3}} \geq \frac{90 - 85}{5/\sqrt{3}}\right) \\
&= \mathbf{P}\left(\frac{S - 85}{5/\sqrt{3}} \geq \sqrt{3}\right) \\
&= 1 - \Phi(\sqrt{3}) \\
&\approx 1 - \Phi(1.73) \\
&\approx 0.04 \\
\mathbf{P}(S' \geq 90) &= \mathbf{P}\left(\frac{S' - \mu_{S'}}{\sigma_{S'}} \geq \frac{90 - \mu_{S'}}{\sigma_{S'}}\right) \\
&= \mathbf{P}\left(\frac{S' - 85}{5/\sqrt{2}} \geq \frac{90 - 85}{5/\sqrt{2}}\right) \\
&= \mathbf{P}\left(\frac{S' - 85}{5/\sqrt{2}} \geq \sqrt{2}\right) \\
&= 1 - \Phi(\sqrt{2}) \\
&\approx 1 - \Phi(1.41) \\
&\approx 0.08 > 0.04
\end{aligned}$$

Thus, Option 2 gives a higher probability of an A than Option 1. However, the probabilities to get a grade lower than a B are:

$$\begin{aligned}
1 - \mathbf{P}(A) - \mathbf{P}(B) &= 1 - 0.04 - 0.92 = 0.04 \text{ under Option 1} \\
1 - \mathbf{P}(A) - \mathbf{P}(B) &= 1 - 0.08 - 0.84 = 0.08 \text{ under Option 2}
\end{aligned}$$

Thus, under Option 2, you are more likely to get an A but also more likely to get a lower grade than B. So your selection of a strategy in this case will depend on your appetite for risk.

Problem 5. Let X and Y be independent random variables, each one uniformly distributed in the interval $[0, 1]$. Find the probability of each of the following events.

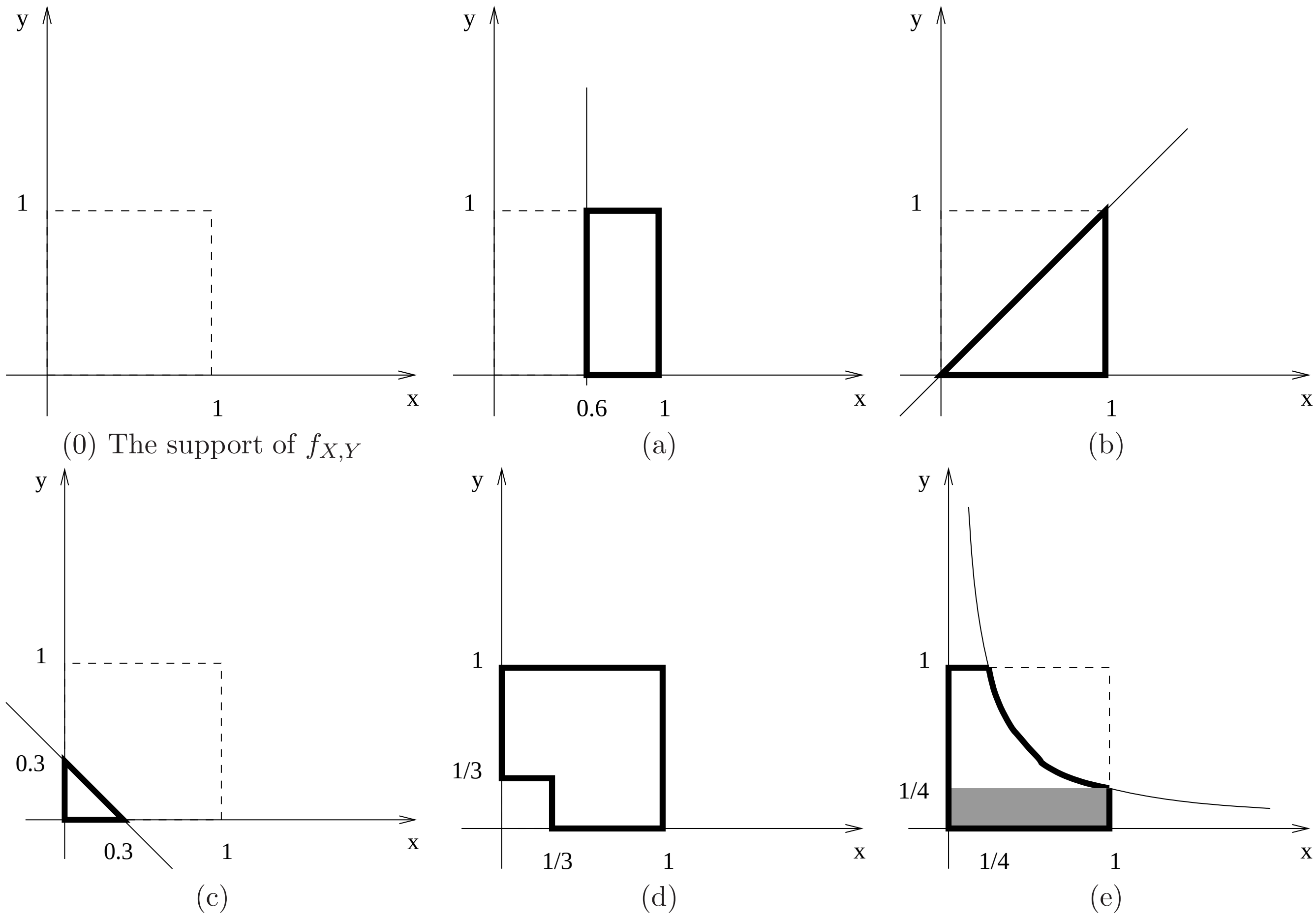


Figure 4: Plots for Problem 5.

(a) $X > 6/10$.

(b) $Y < X$.

(c) $X + Y \leq 3/10$.

(d) $\max\{X, Y\} \geq 1/3$.

(e) $XY \leq 1/4$.

Solution. The joint PDF is equal to one on the unit square depicted in Fig. 4(0). The events for each of the five parts of this problem are depicted in the corresponding parts of Fig. 4. Since the joint PDF is uniform over the unit square, the probability of each event is equal to the area. For parts (a), (b), (c), and (d), these are 0.4, 0.5, $1/2 \cdot 0.3^2 = 0.045$, and $1 - (1/3)^2 = 8/9$, respectively. For part (e), we find the area by integration. To do this, we partition the set into two pieces: the gray piece and

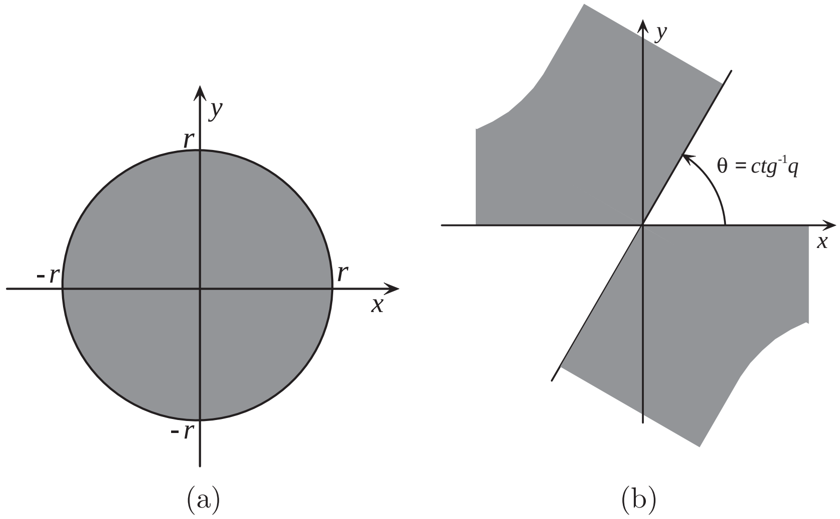


Figure 5: Problem 6.

the rest:

$$\begin{aligned}\mathbf{P}\left(XY \le \frac{1}{4}\right) &= \int_0^{1/4}\int_0^1 dxdy + \int_{1/4}^1\int_0^{1/(4y)} dxdy = \frac{1}{4} + \int_{1/4}^1 \frac{1}{4y}dy \\ &= \frac{1}{4} + \frac{1}{4}\ln y\Big|_{1/4}^1 = \frac{1}{4} - \frac{1}{4}\ln\frac{1}{4} = \frac{1}{4} + \frac{1}{2}\ln 2.\end{aligned}$$

Problem 6. Rayleigh and Cauchy random variables. *(Solutions by Ilya Pollak and Bin Ni.)* A target is located at the origin of a Cartesian coordinate system. One missile is fired at the target, and we assume that X and Y, the coordinates of the missile impact point, are independent random variables each described by the standard normal PDF,

$$f_X(a) = f_Y(a) = \frac{1}{\sqrt{2\pi}}e^{-a^2/2}.$$

(a) Determine the PDF for random variable R, the distance from the target to the point of impact.

Solution. Since R is a distance, we immediately have $f_R(r) = 0$ for $r < 0$. For $r \ge 0$, we have:

$$f_R(r) = \frac{dF_R}{dr}(r) = \frac{d}{dr}\mathbf{P}(R \le r) = \frac{d}{dr}\mathbf{P}(X^2 + Y^2 \le r^2),$$

because $R = \sqrt{X^2 + Y^2}$. This last probability is obtained by integrating the joint PDF of X and Y over the circle of radius r centered at the origin which is depicted in Fig. 5(a). Since X and Y are independent,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}} = \frac{1}{2\pi}e^{-\frac{\rho^2}{2}},$$

where $\rho = \sqrt{x^2 + y^2}$. To integrate this over a circle, it is convenient to use polar coordinates (ρ, θ):

$$\begin{aligned} f_R(r) &= \frac{d}{dr}\mathbf{P}(X^2 + Y^2 \le r^2) = \frac{d}{dr}\int_0^r \int_0^{2\pi} \frac{1}{2\pi} e^{-\frac{\rho^2}{2}} \rho d\theta d\rho = \frac{d}{dr}\int_0^r e^{-\frac{\rho^2}{2}} \rho d\rho \\ &= r e^{-\frac{r^2}{2}}. \end{aligned}$$

Combining the cases $r \ge 0$ and $r < 0$, we obtain:

$$f_R(r) = \begin{cases} r e^{-\frac{r^2}{2}}, & r \ge 0 \\ 0, & r < 0. \end{cases}$$

This distribution is called *Rayleigh* distribution.

(b) Determine the PDF for random variable Q, defined by

$$Q = \frac{X}{Y}.$$

Solution. Let us first find the CDF and then differentiate. By definition,

$$F_Q(q) = \mathbf{P}(Q \le q) = \mathbf{P}(X/Y \le q)$$

We cannot simply multiply both sides of the inequality by Y because we do not know the sign of Y. But we can rewrite the event $\{X/Y \le q\}$ as:

$$\begin{aligned} \{X/Y \le q\} &= (\{X/Y \le q\} \cap \{Y \le 0\}) \cup (\{X/Y \le q\} \cap \{Y > 0\}) \\ &= (\{X \ge qY\} \cap \{Y \le 0\}) \cup (\{X \le qY\} \cap \{Y > 0\}), \end{aligned}$$

which is depicted in Fig. 5(b) as the dark region. The slope of the straight line is $1/q$, and θ is the angle between this line and the x axis. The probability of the event is equal to the integral of $f_{X,Y}$ over the dark region. From Part (a), the joint PDF of X and Y is:

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} = \frac{1}{2\pi} e^{-\frac{\rho^2}{2}}.$$

Because $f_{X,Y}$ is constant on any circle centered at the origin, the integral over the dark region is $(\pi - \theta)/\pi$ of the integral over the entire plane which is 1. Hence we get:

$$\begin{aligned} F_Q(q) &= \mathbf{P}(X/Y \le q) = (\pi - \theta)/\pi \\ &= 1 - \frac{\tan^{-1}(1/q)}{\pi}. \end{aligned}$$

We now differentiate to get:

$$f_Q(q) = \frac{dF_Q(q)}{dq} = \frac{1}{\pi(1 + q^2)}.$$

This distribution is called *Cauchy* distribution. Note that, as $q \to \infty$, the Cauchy PDF decays polynomially (as q^{-2})—much more slowly than normal and exponential PDFs which both decay exponentially. For this reason, Cauchy random variables are used to model phenomena where the probability of outliers (or unusual outcomes) is high. Inconveniently, Cauchy random variables do not have finite means or variances. Other PDFs with polynomial tails, such as Pareto (see Problem 8), are also used to model high probability of outliers.

Problem 7. *(Ilya Pollak.)*
Let X be a continuous random variable, uniformly distributed between -2 and 2, i.e.,

$$f_X(x) = \begin{cases} \frac{1}{4}, & -2 \leq x \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find $E[X]$.

(b) Find $\text{var}(X)$.

(c) Find the PDF of the random variable $Y = X + 1$.

(d) Find the correlation coefficient of $Y = X + 1$ and X.

(e) Find the CDF of X.

(f) Find the CDF of $Z = 2X$ and differentiate to find the PDF of Z.

Solution.

(a) For a continuous random variable which is uniform between a and b, the expected value is $0.5(a + b)$ which in this case is zero.

(b) Since this is a zero-mean random variable, $\text{var}(X) = E[X^2] = \int_{-2}^{2} 0.25x^2 dx = \left.\frac{x^3}{12}\right|_{-2}^{2} = \frac{4}{3}$.

(c) Since X is uniform between -2 and 2, Y is uniform between -1 and 3: $f_Y(y) = 0.25(u(y+1) - u(y-3))$.

(d) Since Y is a linear function of X with a positive slope, the correlation coefficient is equal to one.

(e) The CDF of X is:

$$F_X(x) = \int_{-\infty}^{x} f_X(a)da = \begin{cases} 0, & x < -2 \\ \frac{x+2}{4}, & -2 \leq x \leq 2 \\ 1, & x > 2. \end{cases}$$

(f) The CDF of Z is:

$$\begin{aligned} F_Z(z) &= \mathbf{P}(Z \leq z) = \mathbf{P}(2X \leq z) = \mathbf{P}(X \leq z/2) = F_X(z/2) \\ &= \begin{cases} 0, & z < -4 \\ \frac{z+4}{8}, & -4 \leq z \leq 4 \\ 1, & z > 4. \end{cases} \end{aligned}$$

Differentiating the CDF, we obtain the PDF:

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \begin{cases} \frac{1}{8}, & -4 \leq z \leq 4 \\ 0, & \text{otherwise.} \end{cases}$$

Problem 8. PARETO DISTRIBUTION. *(Ilya Pollak.)*
A *Pareto* random variable X has the following PDF:

$$f_X(x) = \begin{cases} a\frac{c^a}{x^{a+1}}, & \text{for } x \geq c \\ 0, & \text{for } x < c. \end{cases}$$

The two parameters of the PDF, a and c, are positive real numbers.

Pareto random variables are used to model many phenomena in computer science, physics, economics, finance, and other fields.

(a) Find the CDF of X.

(b) Find $E[X]$ for $a > 1$. Your answer will be a function of a and c. Show that $E[X]$ does not exist for $0 < a \leq 1$.

(c) Find $E[X^2]$ for $a > 2$. Your answer will be a function of a and c. Show that $E[X^2]$ does not exist for $0 < a \leq 2$.

(d) Find the variance of X for $a > 2$.

(e) Find the median of X, defined for any continuous random variable as the number m for which $\mathbf{P}(X \leq m) = \mathbf{P}(X \geq m) = 1/2$. Your answer will be a function of a and c.

(f) Let $c = 2$ and $a = 3$. Let μ be the mean of X, and let σ be the standard deviation of X. Compute the probability $\mathbf{P}(X > 3\sigma + \mu)$. Compare this with the probability that a normal random variable is more than three standard deviations above its mean.

Solution.

(a) The CDF is the antiderivative of the PDF. Therefore, for $x < c$, we have that $F_X(x) = 0$. For $x \geq x$,

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(u)du \\ &= \int_{c}^{x} a\frac{c^a}{u^{a+1}}du \\ &= -\frac{c^a}{u^a}\bigg|_c^x du \\ &= 1 - \left(\frac{c}{x}\right)^a \end{aligned}$$

Therefore,

$$F_X(x) = \begin{cases} 1 - \left(\frac{c}{x}\right)^a, & \text{for } x \geq c \\ 0, & \text{for } x < c. \end{cases}$$

(b)

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_c^{\infty} xa\frac{c^a}{x^{a+1}} dx \\ &= \int_c^{\infty} a\frac{c^a}{x^a} dx \end{aligned}$$

First, let's address the case $0 < a \leq 1$. In this case, we have $1/x^a \geq 1/x$ for any $x \geq 1$, and therefore the integrand in the above integral is greater than or equal to $a \cdot c^a/x$ for any $x \geq 1$. Moreover, since the integrand is positive for the entire range of the integral, it follows that reducing the range of integration can only reduce the value of the integral. Putting these considerations together, we have the following lower bound on the value of $E[X]$ when $0 < a \leq 1$:

$$\begin{aligned} E[X] &= \int_c^{\infty} a\frac{c^a}{x^a} dx \\ &\geq \int_{\max(c,1)}^{\infty} a\frac{c^a}{x^a} dx \\ &\geq \int_{\max(c,1)}^{\infty} a\frac{c^a}{x} dx \\ &= ac^a \int_{\max(c,1)}^{\infty} \frac{1}{x} dx \\ &= ac^a \ln x|_{\max(c,1)}^{\infty}, \end{aligned}$$

which diverges. Since the last expression is a lower bound for the original integral, the original integral also diverges. Therefore, $E[X]$ does not exist for $0 < a \leq 1$.

For the case $a > 1$, we have:

$$\begin{aligned} E[X] &= \int_c^{\infty} a\frac{c^a}{x^a} dx \\ &= -a\frac{c^a}{(a-1)x^{a-1}}\bigg|_c^{\infty} \\ &= \frac{ac}{a-1} \end{aligned}$$

(c) Using the definition of expectation,

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\ &= \int_c^{\infty} x^2 a\frac{c^a}{x^{a+1}} dx \\ &= \int_c^{\infty} a\frac{c^a}{x^{a-1}} dx \end{aligned}$$

When $0 < a \leq 2$, an argument similar to Part (b) shows that the integral diverges and therefore $E[X^2]$ does not exist. For $a > 2$,

$$\begin{aligned}
E[X^2] &= \int_c^\infty a\frac{c^a}{x^{a-1}}dx \\
&= -a\frac{c^a}{(a-2)x^{a-2}}\Big|_c^\infty \\
&= \frac{ac^2}{a-2}
\end{aligned}$$

(d) Using the first two moments of X found in Parts (b) and (c), we have:

$$\begin{aligned}
\text{var}(X) &= E[X^2] - (E[X])^2 \\
&= \frac{ac^2}{a-2} - \frac{a^2c^2}{(a-1)^2} \\
&= \frac{ac^2(a-1)^2 - a^2c^2(a-2)}{(a-2)(a-1)^2} \\
&= \frac{ac^2[(a-1)^2 - a(a-2)]}{(a-2)(a-1)^2} \\
&= \frac{ac^2}{(a-2)(a-1)^2}
\end{aligned}$$

(e) To find the median, we take the CDF found in Part (a) and set it equal to 1/2:

$$\begin{aligned}
1 - \left(\frac{c}{\text{median}[X]}\right)^a &= 1/2 \\
\frac{c}{\text{median}[X]} &= (1/2)^{1/a} \\
\text{median}[X] &= c \cdot 2^{1/a}
\end{aligned}$$

(f) From part (b), the mean is 3. From part (d), the variance is $(2^2 \cdot 3)/(2^2 \cdot 1) = 3$, and therefore the standard deviation is $\sqrt{3}$. Hence, the $3\sigma + \mu = 3 + 3\sqrt{3}$, and

$$\begin{aligned}
\mathbf{P}(X > 3\sigma + \mu) &= 1 - F_X(3 + 3\sqrt{3}) \\
&= \left(\frac{2}{3 + 3\sqrt{3}}\right)^3 \\
&= \frac{8}{27(1+\sqrt{3})^3} \\
&= \frac{8}{27(1 + 3\sqrt{3} + 3 \cdot 3 + 3\sqrt{3})} \\
&= \frac{8}{27(10 + 6\sqrt{3})} \\
&= \frac{4}{27(5 + 3\sqrt{3})} \\
&\approx 0.014530.
\end{aligned}$$

Checking the table of normal CDF values, we see that the probability for a normal random variable to be three standard deviations or more above the mean is 0.001350—i.e., more than a factor of 10 lower than the probability 0.014530 we obtained above for a Pareto random variable. For Pareto random variables, outliers are much more likely than for normal random variables, because a Pareto PDF decays polynomially for large outcomes whereas a normal PDF decays exponentially. Many physical phenomena result in observations which are quite likely to be more than three standard deviations away from the mean—for example, the returns of financial instruments. For such observations, normal models are inappropriate, and models with polynomial tails, such as Pareto, are often used instead.

Problem 9. DEPENDENCE AND CORRELATEDNESS. *(Ilya Pollak.)*
This problem reviews the following fact: independence implies uncorrelatedness but uncorrelatedness does not imply independence. Equivalently, correlatedness implies dependence, but dependence does not imply correlatedness. Said another way, two independent random variables are always uncorrelated. Two correlated random variables are always dependent. However, two uncorrelated random variables are not necessarily independent; and two dependent random variables are not necessarily correlated.

(a) Continuous random variables X and Y are independent. Prove that X and Y are uncorrelated.

Solution. Independence of X and Y means that $f_{XY}(x,y) = f_X(x)f_Y(y)$. Therefore,

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)(y)dxdy \\ &= \int_{-\infty}^{\infty} xf_X(x)dx\int_{-\infty}^{\infty} yf_Y(y)(y)dy \\ &= E[X]E[Y], \end{aligned}$$

which means that X and Y are uncorrelated.

(b) Discrete random variables X and Y are independent. Prove that X and Y are uncorrelated.

Solution. Independence of X and Y means that $p_{XY}(x,y) = p_X(x)p_Y(y)$. Therefore,

$$\begin{aligned} E[XY] &= \sum_{x,y} xyp_{XY}(x,y) \\ &= \sum_{x,y} xyp_X(x)p_Y(y)(y) \\ &= \sum_{x} xp_X(x)\sum_{y} yp_Y(y)(y) \\ &= E[X]E[Y], \end{aligned}$$

which means that X and Y are uncorrelated.

(c) Continuous random variables X and Y are correlated. Prove that they are not independent.

Solution. If X and Y were independent, they would be uncorrelated, according to Part (a). Thus, if they are correlated, they cannot be independent.

(d) Discrete random variables X and Y are correlated. Prove that they are not independent.

Solution. If X and Y were independent, they would be uncorrelated, according to Part (b). Thus, if they are correlated, they cannot be independent.

(e) Suppose X is a discrete random variable, uniformly distributed between -1 and 1, and let $Y = X^2$. Show that X and Y are uncorrelated but not independent. Also show that $E[Y|X=0] \neq E[Y]$.

Solution. $E[X] = 0$. The mean of Y is:

$$E[Y] = E[X^2] = 1 \cdot (1/3) + 0 \cdot (1/3) + 1 \cdot (1/3) = 2/3.$$

The mean of XY is:

$$E[XY] = E[X^3] = -1 \cdot (1/3) + 0 \cdot (1/3) + 1 \cdot (1/3) = 0 = E[X]E[Y],$$

and therefore X and Y are uncorrelated. The marginal distribution of X is:

$$p_X(x) = \begin{cases} 1/3, & \text{for } x = -1, 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

The marginal distribution of Y is:

$$p_Y(y) = \begin{cases} 1/3, & \text{for } y = 0 \\ 2/3, & \text{for } y = 1 \\ 0, & \text{otherwise} \end{cases}$$

The joint distribution of X and Y is:

$$p_{X,Y}(x,y) = \begin{cases} 1/3, & \text{for } x = -1 \text{ and } y = 1 \\ 1/3, & \text{for } x = 0 \text{ and } y = 0 \\ 1/3, & \text{for } x = 1 \text{ and } y = 1 \\ 0, & \text{otherwise} \end{cases}$$

The joint PMF is not equal to the product of the marginal PMF's: for example, $p_{X,Y}(0,0) = 1/3 \neq p_X(0)p_Y(0) = 1/9$. Therefore, the two random variables are not independent.

Given $X = 0$, we have that $Y = 0$ with probability 1. Therefore, $E[Y|X=0] = 0 \neq E[Y] = 2/3$.

Problem 10. A UNIFORM JOINT DISTRIBUTION OF TWO CONTINUOUS RANDOM VARIABLES. *(Ilya Pollak.)*
Consider the following joint PDF of two continuous random variables, X and Y:

$$f_{XY}(x,y) = \begin{cases} A, & |x| + |y| \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find A.

(b) Find the marginal PDFs $f_X(x)$ and $f_Y(y)$.

(c) Are X and Y uncorrelated? Are they independent?

(d) Find $E[Y]$ and $\text{var}(Y)$.

(e) Find the conditional PDF $f_{Y|X}(y|0.5)$. In other words, find the conditional PDF of Y given that $X = 0.5$. Find the corresponding conditional CDF $F_{Y|X}(y|0.5)$.

Solution.

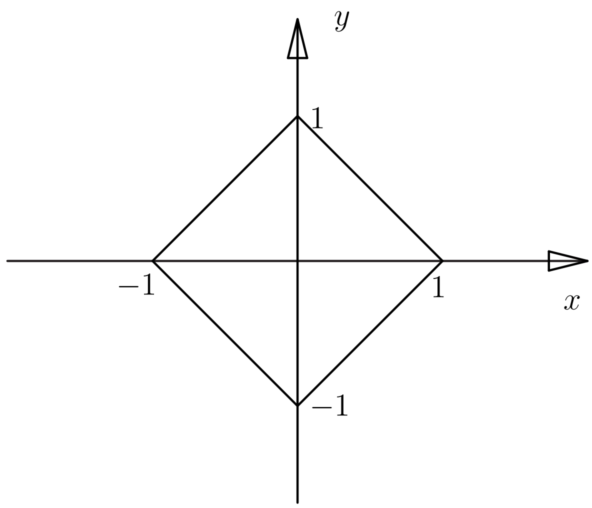


Figure 6: $|x| + |y| \leq 1$

(a) The region described by $|x| + |y| \leq 1$ is shown in Figure 6 and its area is 2. To find A,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy \\ &= A \cdot 2 \\ \Rightarrow A &= \frac{1}{2}. \end{aligned}$$

(b) Let us calculate $f_X(x)$ in the following two regions: $-1 \leq x \leq 0$ and $0 < x \leq 1$.
When $-1 \leq x \leq 0$:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_{-1-x}^{1+x} Ady = A \cdot [(1+x) - (-1-x)] = 1 + x$$

When $0 < x \leq 1$:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_{x-1}^{1-x} Ady = A \cdot [(1-x) - (x-1)] = 1 - x$$

Therefore,

$$f_X(x) = \begin{cases} 1-|x|, & -1 \le x \le 1 \\ 0, & \text{else} \end{cases}$$

Similarly, we can find

$$f_Y(y) = \begin{cases} 1-|y|, & -1 \le y \le 1 \\ 0, & \text{else} \end{cases}$$

(c) Since $f_X(x)$ is an even function, $xf_X(x)$ is an odd function. The integral of $xf_X(x)$ from $-\infty$ to ∞ is therefore zero. Therefore, $E[X] = 0$. Similarly, $E[Y] = 0$. The function $f_{XY}(x, y)$ is even in both x and y, and therefore $xyf_{XY}(x, y)$ is odd in both x and y. This means that the integral of $xyf_{XY}(x, y)$ over the whole plane is zero, i.e., $E[XY] = 0 = E[X]E[Y]$, and therefore X and Y are uncorrelated. However, X and Y are not independent because $f_X(x) \cdot f_Y(y) \neq f_{XY}(x, y)$. For example, for the point $x = y = 0$, we have $f_X(0) = f_Y(0) = 1$ and therefore $f_X(0) \cdot f_Y(0) = 1$ whereas $f_{XY}(0, 0) = 1/2$.

(d) From Part (c), $E[Y] = 0$. Therefore,

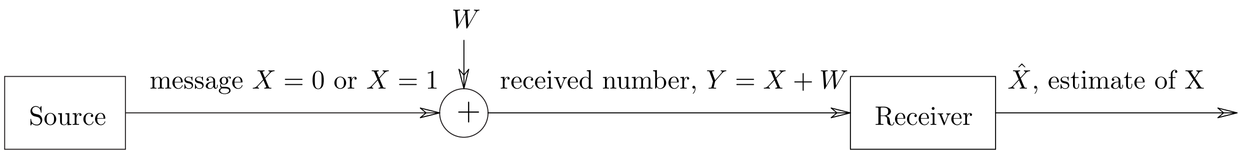
$$\begin{aligned} \text{var}(Y) &= E[Y^2] = \int_{-\infty}^{\infty} y^2 \cdot f_Y(y)dy = \int_{-1}^{0} y^2 \cdot (1+y)dy + \int_{0}^{1} y^2 \cdot (1-y)dy \\ &= \int_{-1}^{0} (y^2 + y^3)dy + \int_{0}^{1} (y^2 - y^3)dy \\ &= \left(\frac{y^3}{3} + \frac{y^4}{4}\right)\Bigg|_{-1}^{0} + \left(\frac{y^3}{3} - \frac{y^4}{4}\right)\Bigg|_{0}^{1} \\ &= \frac{1}{6} \end{aligned}$$

(e) Using the definition of conditional PDF,

$$\begin{aligned} f_{Y|X}(y|0.5) &= \frac{f_{XY}(0.5, y)}{f_X(0.5)} \\ &= \begin{cases} \frac{0.5}{1-|0.5|}, & |0.5| + |y| \le 1 \\ \frac{0}{1-|0.5|}, & \text{otherwise} \end{cases} \\ &= \begin{cases} 1, & |y| \le 0.5 \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

i.e., the conditional density is uniform between $-1/2$ and $1/2$. The conditional CDF is the antiderivative of the conditional PDF:

$$F_{Y|X}(y|0.5) = \begin{cases} 0, & y \le -0.5 \\ y + 0.5, & |y| \le 0.5 \\ 1, & y \ge 0.5 \end{cases}$$



Problem 11. (Minimum error probability detection.) *(Ilya Pollak.)*

In the binary communication system shown above, messages $X = 0$ and $X = 1$ occur with probabilities p_0 and $p_1 = 1 - p_0$, respectively. The random variable W is normal (Gaussian), with mean zero and standard deviation σ. The random variables X and W are independent. The receiver sees one observation of the random variable $Y = X + W$, and attempts to detect X–i.e., for every possible observed value Y it chooses an estimate $\hat{X}(Y)$. Consider the following detector of X:

$$\hat{X}(Y) = \begin{cases} 1 & \text{if } Y \geq \eta, \\ 0 & \text{if } Y < \eta, \end{cases} \tag{3}$$

where the decision threshold η is a real number. In other words, if the observed value of Y is $\geq \eta$, then we decide that 1 was transmitted; otherwise, we decide that 0 was transmitted.

(a) Find an expression for $\mathbf{P}(\hat{X}(Y) = X | X = 0)$. This is the conditional probability of correct decision, given than 0 was transmitted. Your expression will be an integral of a probability density, and will depend on the parameters η and σ.

(b) Find a similar expression for $\mathbf{P}(\hat{X}(Y) = X | X = 1)$, the conditional probability of correct decision given than 1 was transmitted.

(c) Apply the total probability theorem to your results from Parts (a) and (b) to find an expression for the probability of correct decision, $\mathbf{P}(\hat{X}(Y) = X)$.

(d) Find the decision threshold η which achieves the maximum probability of correct decision. In other words, find η which maximizes $\mathbf{P}(\hat{X}(Y) = X)$. To do this, differentiate with respect to η the expression for $\mathbf{P}(\hat{X}(Y) = X)$ you found in Part (c), set to zero, and solve for η. Your answer will be an expression involving σ and p_0. Then show that the answer you got for η is a maximum.

(e) Show that the receiver of Eq. (3), with the threshold η you obtained in Part (d), is equivalent to the following rule:
if $f_{Y|X}(y|1)p_1 \geq f_{Y|X}(y|0)p_0$, decide that 1 was transmitted; otherwise, decide that 0 was transmitted.

(f) Show that your receiver is also equivalent to the following rule:
if $\mathbf{P}(X = 1|Y = y) \geq \mathbf{P}(X = 0|Y = y)$, decide that 1 was transmitted; otherwise, decide that 0 was transmitted.
This detector is called the *maximum aposteriori probability* detector, since it chooses the hypothesis whose conditional probability given the observed data is maximum. You can use the

following identities:

$$\begin{aligned}\mathbf{P}(X=1|Y=y) &= \frac{f_{Y|X}(y|1)p_1}{f_Y(y)},\\ \mathbf{P}(X=0|Y=y) &= \frac{f_{Y|X}(y|0)p_0}{f_Y(y)}.\end{aligned}$$

(g) Fix $p_0 = 0.5$. What is the corresponding value of η? Sketch $f_{Y|X}(y|1)p_1$ and $f_{Y|X}(y|0)p_0$ for:

(i) $\sigma = 0.1$;

(ii) $\sigma = 0.5$;

(iii) $\sigma = 1$.

In each case, use the same coordinate axes for both functions. From looking at the plots, in which of the three cases would you expect the probability of correct decision to be the largest? the smallest? Use a table of the normal CDF to find the probability of correct decision in each of the three cases. What will happen to the probability of correct decision as $\sigma \to \infty$? as $\sigma \to 0$?

(h) Now suppose $\sigma = 0.5$ and $p_0 \approx 1$ (say, e.g., $p_0 = 0.9999$). What is, approximately, the probability of correct decision? What is, approximately, the conditional probability $\mathbf{P}(\hat{X}(Y) = X|X = 1)$? Suppose now that the message "$X = 0$" is "your house is not on fire" and the message "$X = 1$", which has a very low probability, is "your house is on fire". It may therefore be much more valuable to correctly detect "$X = 1$" than "$X = 0$". Suppose we get -\$10 if we decide that the house is on fire when in fact it is, but that we get -\$1000 if we decide that the house is not on fire when in fact it is. Suppose further that we get \$0 if we correctly decide that the house is not on fire, and that we get -\$1 if we incorrectly decide that the house is on fire. It would make sense to maximize something other than the probability of correct decision. Propose such a criterion. (You do not have to find the optimal solution for the new criterion.)

Solution.

(a) When $X = 0$, $Y = W$. Hence we have:

$$f_{Y|X}(y|0) = f_W(y) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y^2}{2\sigma^2}},$$

$$\mathbf{P}(\hat{X}(Y) = X|X = 0) = \mathbf{P}(Y < \eta|X = 0) = \int_{-\infty}^{\eta} f_{Y|X}(y|0)dy = \int_{-\infty}^{\eta} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y^2}{2\sigma^2}}dy.$$

(b) When $X = 1$, $Y = 1 + W$. Hence we have:

$$f_{Y|X}(y|1) = f_W(y-1) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-1)^2}{2\sigma^2}},$$

$$\mathbf{P}(\hat{X}(Y) = X|X = 1) = \mathbf{P}(Y >= \eta|X = 1) = \int_{\eta}^{\infty} f_{Y|X}(y|1)dy = \int_{\eta}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-1)^2}{2\sigma^2}}dy.$$

(c) By the total probability theorem:

$$\begin{aligned}\mathbf{P}(\hat{X}(Y)=X) &= \mathbf{P}(\hat{X}(Y)=X|X=0)\mathbf{P}(X=0)+\mathbf{P}(\hat{X}(Y)=X|X=1)\mathbf{P}(X=1)\\ &= p_0\int_{-\infty}^{\eta} f_{Y|X}(y|0)dy + p_1\int_{\eta}^{\infty} f_{Y|X}(y|1)dy. \qquad (4)\end{aligned}$$

(d) For notational convenience, let us call $\phi(\eta)=\mathbf{P}(\hat{X}(Y)=X)$. Taking the derivative of Eq. (4) with respect to η and equating it to zero, we get:

$$\begin{aligned}\phi'(\eta) &= \frac{d\mathbf{P}(\hat{X}(Y)=X)}{d\eta} = p_0 f_{Y|X}(\eta|0) - p_1 f_{Y|X}(\eta|1) \qquad (5)\\ &= p_0\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{\eta^2}{2\sigma^2}\right] - p_1\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\eta-1)^2}{2\sigma^2}\right]=0 \qquad (6)\\ \Rightarrow \frac{p_0}{p_1} &= \exp\left[\frac{\eta^2-(\eta-1)^2)}{2\sigma^2}\right] = \exp\left[\frac{2\eta-1}{2\sigma^2}\right]\\ \Rightarrow \eta_{max} &= \sigma^2\ln\frac{p_0}{p_1}+\frac{1}{2}. \qquad (7)\end{aligned}$$

Thus, η_{max} is the unique extremum of $\phi(\eta)$. We now need to show that it's a maximum. Note that, if $p_0 > p_1$, then $\eta_{max} > 1/2$. From the left side of Eq. (6), $\phi'(1/2) > 0$–i.e., the function is increasing on the left of η_{max}, which means that η_{max} is a maximum. Similarly, if $p_0 < p_1$, then $\eta_{max} < 1/2$, and $\phi'(1/2) < 0$, and so the function is decreasing on the right of η_{max}, which means that η_{max} is again a maximum. When $p_0 = 1/2$, we infer from Eq. (4) that $\phi(\infty)=\phi(-\infty)=1/2$ whereas $\phi(\eta_{max})=\phi(1/2)=\Phi(1/(2\sigma))>1/2$, which means that again, η_{max} is a maximum.

(e) In Part (d), we showed that η_{max} is the unique maximum of the function $\phi(\eta)$. This means that, for $y<\eta_{max}$, $\phi'(y)>0$, and for $y\geq\eta_{max}$, $\phi'(y)\leq 0$. Using the expression for $\phi'(y)$ we obtained in Eq. (5), we see that this is equivalent to: if $Y=y\geq\eta_{max}$ (i.e. we decide that 1 was sent), then $p_0 f_{Y|X}(y|0)\leq p_1 f_{Y|X}(y|1)$; if $Y=y<\eta_{max}$ (we decide that 0 was sent), then $p_0 f_{Y|X}(y|0) > p_1 f_{Y|X}(y|1)$. This is illustrated in Fig. 7.

(f) Use

$$\begin{aligned}\mathbf{P}(X=1|Y=y) &= \frac{f_{Y|X}(y|1)p_1}{f_Y(y)},\\ \mathbf{P}(X=0|Y=y) &= \frac{f_{Y|X}(y|0)p_0}{f_Y(y)}.\end{aligned}$$

Combining these two equations with Part (e), we see that:

$$\begin{aligned}\mathbf{P}(X=1|Y=y)\geq\mathbf{P}(X=0|Y=y) &\Leftrightarrow p_1 f_{Y|X}(y|1)\geq p_0 f_{Y|X}(y|0)\\ &\Leftrightarrow \text{decide that 1 was sent.}\end{aligned}$$

(g) Substituting $p_0 = 0.5$ in Eq. (7), we immediately get:

$$\eta_{max}=\sigma^2\ln\frac{0.5}{1-0.5}+\frac{1}{2}=\frac{1}{2}.$$

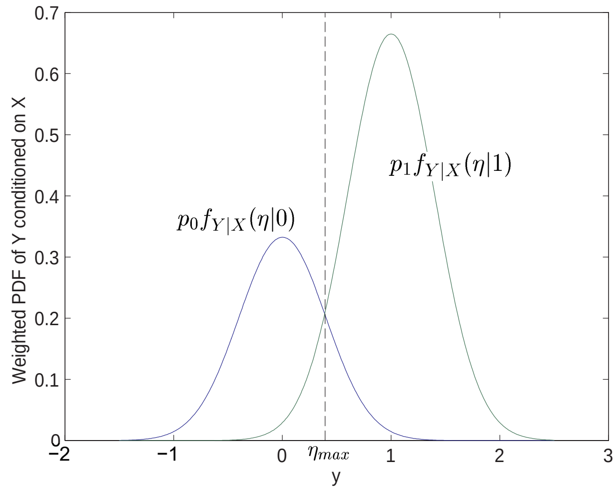
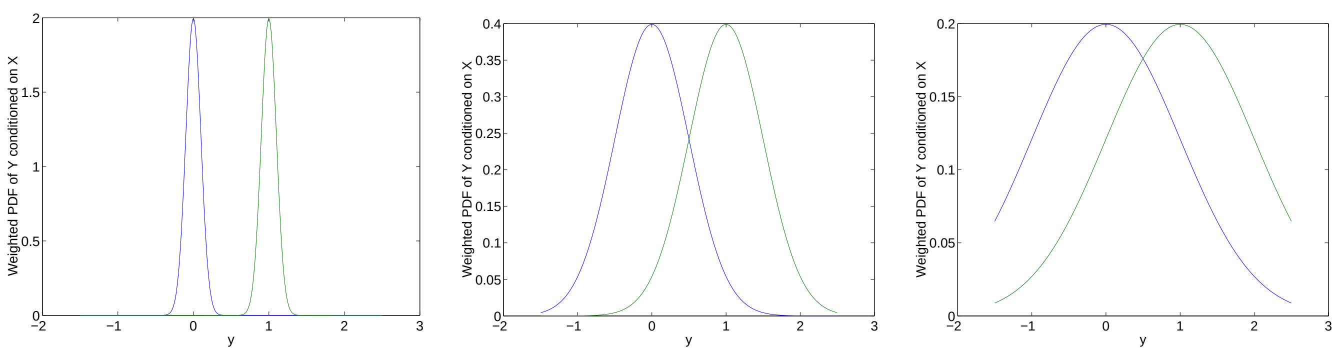


Figure 7: Weighted PDF's of Y conditioned on $X = 0$ and on $X = 1$.



(a) $\sigma = 0.1$ (b) $\sigma = 0.5$ (c) $\sigma = 1$

Figure 8: Weighted PDF's of Y conditioned on X.

(i,ii,iii) $p_1 f_{Y|X}(y|1)$ and $p_0 f_{Y|X}(y|0)$ for all the three cases are plotted in Fig. 8. Case (i) has the largest probability of correct decision: among the three plots, this is the one where the smallest portions of the conditional PDF's "spill over" to the wrong side of the threshold $\eta_{max} = 1/2$.

The probability of correct decision can be determined from Eq. (4):

$$\begin{aligned}
\mathbf{P}(\hat{X}(Y) = X) &= \frac{1}{2}\int_{-\infty}^{\eta_{max}} f_{Y|X}(y|0)dy + \frac{1}{2}\int_{\eta_{max}}^{\infty} f_{Y|X}(y|1)dy \\
&= \frac{1}{2}\Phi\left(\frac{\eta_{max}}{\sigma}\right) + \frac{1}{2}\left[1 - \Phi\left(\frac{\eta_{max} - 1}{\sigma}\right)\right] \\
&= \frac{1}{2}\Phi\left(\frac{1}{2\sigma}\right) + \frac{1}{2}\left[1 - \Phi\left(-\frac{1}{2\sigma}\right)\right] \\
&= \Phi\left(\frac{1}{2\sigma}\right).
\end{aligned}$$

The probabilities of correct decision for the three cases are 0.9999997, 0.8413447, 0.6914624 respectively.

When the noise is extremely large, it becomes impossible to distinguish between $X = 0$ and $X = 1$; therefore, when $\sigma \to \infty$, both the probability of correct decision and the error probability tend to 1/2. When $\sigma \to 0$, the conditional PDF's become very well concentrated around their means; there is no noise in this case, and $Y \to X$. The probability of correct decision will tend to 1.

(h) When p_0 is approximately 1, p_1 is approximately 0. From Eq. (7), and also from Fig. 7, we can see that η_{max} goes to ∞ in this case. Therefore Eq. (4) becomes:

$$\mathbf{P}(\hat{X}(Y) = X) \approx \int_{-\infty}^{\infty} f_{Y|X}(y|0)dy = 1.$$

So we have probability 1 to make the correct decision: since it's virtually certain that $X = 0$ is transmitted, the receiver just guesses $\hat{X} = 0$ most of the time, and is correct most of the time. However, the *conditional* probability of correct decision given $X = 1$ is:

$$\begin{aligned}\mathbf{P}(\hat{X}(Y) = X|X = 1\} &= \int_{\eta_{max}}^{\infty} f_{Y|X}(y|1)dy \\ &= \int_{\infty}^{\infty} f_{Y|X}(y|1)dy \approx 0.\end{aligned}$$

If we decide $\hat{X} = 0$ all the time, we are bound to make an error whenever $X = 1$ is transmitted. Thus, even though the overall probability of error is miniscule, the conditional probability of error given $X = 1$ is huge. Therefore, if it is very costly to us to make an error when $X = 1$ (when, e.g., "X=1" is "your house is on fire"), it makes sense to try to minimize the *expected cost* (rather than simply the probability of error)–i.e. the expected value of loss due to fire and false alarms:

$$\begin{aligned}\min_{\text{all estimators } \hat{X}(Y)} (10 \cdot \mathbf{P}(\hat{X} = 1 \text{ and } X = 1) &+ 1000 \cdot \mathbf{P}(\hat{X} = 0 \text{ and } X = 1) \\ &+ 1 \cdot \mathbf{P}(\hat{X} = 1 \text{ and } X = 0)).\end{aligned}$$

Problem 12. GAUSSIAN MIXTURE.

A signal $s = 3$ is transmitted from a satellite but is corrupted by noise, and the received signal is $X = s + W$. When the weather is good, which happens with probability 2/3, W is a normal (Gaussian) random variable with zero mean and variance 4. When the weather is bad, W is normal with zero mean and variance 9. In the absence of any weather information:

(a) What is the PDF of X? (**Hint.** Use the total probability theorem.)

(b) Calculate the probability that X is between 2 and 4.

Solution.

(a) What we can do is to first find the CDF of X and then take its derivative to find the PDF. By definition, we have:

$$F_X(x) = \mathbf{P}(X \leq x).$$

Since we don't have any weather information, if we can calculate the above probability conditioned on all the weather situations, we can apply the total probability theorem to get the final answer. If the weather is good, the random variable $X = 3 + W$ is conditionally normal with mean $\mu = 3$ and variance $\sigma^2 = 4$. In this case, the conditional probability that X is less than or equal to some x is:

$$P(X \leq x|good) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-3}{2}\right)$$

Similarly, if we suppose the weather is bad, the conditional probability is:

$$P(X \leq x|bad) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-3}{3}\right)$$

Apply the total probability theorem to get:

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(X \leq x|good)P(good) + P(X \leq x|bad)P(bad) \\ &= \Phi\left(\frac{x-3}{2}\right)\frac{2}{3} + \Phi\left(\frac{x-3}{3}\right)\frac{1}{3} \end{aligned}$$

Finally,

$$\begin{aligned} f_X(x) &= \frac{dF_X(x)}{dx} = \frac{2}{3}\frac{1}{\sqrt{2\pi}2}e^{-\frac{(x-3)^2}{2\cdot 2^2}} + \frac{1}{3}\frac{1}{\sqrt{2\pi}3}e^{-\frac{(x-3)^2}{2\cdot 3^2}} \\ &= \frac{1}{3}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-3)^2}{8}} + \frac{1}{9}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-3)^2}{18}} \end{aligned}$$

In general, the total probability theorem implies that, if $B_1, B_2, \cdots, B_n$ partition the sample space, the following is true for any continuous random variable X:

$$f_X(x) = f_{X|B_1}(x)P(B_1) + f_{X|B_2}(x)P(B_2) + \cdots + f_{X|B_n}(x)P(B_n).$$

(b) To calculate $P(2 \leq X \leq 4)$, we use our results from Part (a):

$$\begin{aligned} P(2 \leq X \leq 4) &= P(X \leq 4) - P(X \leq 2) = F_X(4) - F_X(2) \\ &= \Phi\left(\frac{4-3}{2}\right)\frac{2}{3} + \Phi\left(\frac{4-3}{3}\right)\frac{1}{3} - \left[\Phi\left(\frac{2-3}{2}\right)\frac{2}{3} + \Phi\left(\frac{2-3}{3}\right)\frac{1}{3}\right] \\ &= \Phi\left(\frac{1}{2}\right)\frac{2}{3} + \Phi\left(\frac{1}{3}\right)\frac{1}{3} - \left[\Phi\left(\frac{-1}{2}\right)\frac{2}{3} + \Phi\left(\frac{-1}{3}\right)\frac{1}{3}\right] \\ &\approx \frac{2}{3}\cdot 2 \cdot 0.19146 + \frac{1}{3}\cdot 2\cdot 0.12930 = 0.3415. \end{aligned}$$

Problem 13. *(Ilya Pollak.)*
Consider the following joint PDF of two continuous random variables, X and Y:

$$f_{XY}(x,y) = \begin{cases} A, & 0 \leq x \leq 2 \quad \text{and} \quad 0 \leq y \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find A.

(b) Find the marginal PDFs $f_X(x)$ and $f_Y(y)$.

(c) Are X and Y uncorrelated? Are they independent?

(d) Find $E[Y]$ and $Var(Y)$.

(e) Let $Z = X + Y$. Find the conditional PDF $f_{Z|X}(z|x)$, as well as the marginal PDF $f_Z(z)$ of Z.

Solution. Note that $f_{XY}(x, y)$ can be written as $A(u(x) - u(x-2))(u(y) - u(y-2))$ for all $x, y \in \mathbb{R}$.

(a) The integral of $f_{XY}(x, y)$ over the whole real plane must be 1:

$$\begin{aligned}\int_{\mathbb{R}^2} f_{XY}(x,y)dxdy &= \int_{\mathbb{R}^2} A(u(x) - u(x-2))(u(y) - u(y-2))dxdy \\ &= A\int_0^2\int_0^2 1dxdy \\ &= A \cdot 4 = 1.\end{aligned}$$

Therefore we get $A = 1/4$.

(b) To get the marginal density of X, integrate the joint density:

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy \\ &= \int_{-\infty}^{\infty} 0.25(u(x) - u(x-2))(u(y) - u(y-2))dy \\ &= 0.25(u(x) - u(x-2))\int_{-\infty}^{\infty}(u(y) - u(y-2))dy \\ &= 0.25(u(x) - u(x-2))\int_0^2 1\,dy \\ &= 0.5(u(x) - u(x-2)).\end{aligned}$$

So X is uniformly distributed on [0,2].

$$\begin{aligned}f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dx \\ &= 0.25(u(y) - u(y-2))\int_{-\infty}^{\infty}(u(x) - u(x-2))dx \\ &= 0.5(u(y) - u(y-2)).\end{aligned}$$

Y is also uniformly distributed on [0,2].

(c) From Parts (a) and (b), we have $f_{XY}(x, y) = f_X(x)f_Y(y)$. By definition of independence, X and Y are independent, and therefore also uncorrelated.

(d) By definition of expectation and variance,

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= \int_0^2 0.5y dy = 1. \end{aligned}$$

$$\begin{aligned} Var(Y) &= E[(Y - E[Y])^2] = E[Y^2] - (E[Y])^2 \\ &= \int_0^2 0.5y^2 dy - 1 \\ &= 0.5(1/3)y^3\big|_0^2 - 1 \\ &= 4/3 - 1 = 1/3. \end{aligned}$$

(e) When $X = x$, $Z = x + Y$. Therefore,

$$\begin{aligned} F_{Z|X}(z|x) &= \mathbf{P}(Z \le z | X = x) \\ &= \mathbf{P}(Y \le z - x | X = x) \\ &= \mathbf{P}(Y \le z - x) \qquad \text{(because } X \text{ and } Y \text{ are independent)} \\ &= F_Y(z - x). \end{aligned}$$

Differentiating with respect to z, we get:

$$f_{Z|X}(z|x) = \frac{\partial F_{Z|X}(z|x)}{\partial z} = f_Y(z - x) = 0.5(u(z - x) - u(z - x - 2)).$$

The PDF of Z can be obtained as follows:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{XZ}(x, z) dx = \int_{-\infty}^{\infty} f_{Z|X}(z|x) f_X(x) dx \\ &= \int_{-\infty}^{\infty} f_Y(z - x) f_X(x) dx = f_Y * f_X(z) \\ &= \int_{-\infty}^{\infty} 0.5(u(z - x) - u(z - x - 2)) 0.5(u(x) - u(x - 2)) dx \\ &= \int_0^2 0.25(u(z - x) - u(z - x - 2)) dx \\ &= \begin{cases} z/4, & 0 \le z \le 2 \\ 1 - z/4, & 2 < z \le 4 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Problem 14. Scalar vs vector quantization. *(Ilya Pollak.)*

Many physical quantities can assume an infinite number of values. In order to store them digitally, they need to be approximated using a finite number of values. If the approximated quantity can have r possible values, it can be stored using $\lceil \log_2 r \rceil$ bits, where $\lceil x \rceil$ denotes the smallest integer not exceeding x.

For example, in the physical world, an infinite number of shades of different colors is possible; however capturing a picture with a digital camera and storing it requires (among other things!) approximating the color intensities using a finite range of values, for example, 256 distinct values per color channel. Audio signals can similarly have an infinite range of values; however, in order to record music to an iPod or a CD, a similar approximation is required.

In addition, most of real-world data storage and transmission applications involve lossy data compression—i.e., the use of algorithms that can encode data at different quality levels, through a trade-off between quality and file size. Widely used examples of such algorithms are JPEG (for still pictures), H.264 (for video), and MP3 (for audio). All such lossy data compression schemes involve approximating variables that can assume 2^k values with variables that can only assume 2^n values, with $n < k$.

This process of approximating a variable X that can assume more than 2^n different values, with another variable Y that can only assume 2^n different values, is called *n-bit quantization*.

Suppose A_1 and A_2 are two random variables with the following joint PDF:

$$f_{A_1,A_2}(a_1, a_2) = \begin{cases} \frac{1}{2}, & \text{if both } 0 \le a_1 \le 1 \text{ and } 0 \le a_2 \le 1; \\ \frac{1}{2}, & \text{if both } 1 \le a_1 \le 2 \text{ and } 1 \le a_2 \le 2; \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the marginal PDF of A_1, $f_{A_1}(a_1)$.

(b) ONE-BIT LLOYD-MAX QUANTIZER. Suppose we want to quantize A_1 to one bit, in the following way. We pick three real numbers x_0, q_1, and q_2. If $A_1 < x_0$ we quantize it to q_1. If $A_1 \ge x_0$ we quantize it to q_2. In other words, the quantized version $\hat{A}_1$ of A_1 is determined as follows:

$$\hat{A}_1 = \begin{cases} q_1, & \text{if } A_1 < x_0 \\ q_2, & \text{if } A_1 \ge x_0 \end{cases}$$

Note that, since $\hat{A}_1$ can only assume one of two values, any observation of $\hat{A}_1$ can be stored using one bit. Unfortunately, $\hat{A}_1$ will be only an approximation of A_1. Find the numbers x_0, q_1, and q_2 which minimize the mean-square error (MSE),

$$E[(\hat{A}_1 - A_1)^2]. \tag{8}$$

It can be shown that $0 \le q_1 \le x_0 \le q_2 \le 2$. You can just assume that these inequalities hold, without proof.

(c) Calculate the mean-square error defined in Eq. (8) for the quantizer that you obtained in Part (b).

(d) Show that the marginal PDF of A_2 is the same as the marginal PDF of A_1. Therefore, the one-bit Lloyd-Max quantizer for A_2 is identical to the one for A_1 which you obtained in Part (b), and

$$E[(\hat{A}_2 - A_2)^2] = E[(\hat{A}_1 - A_1)^2]. \tag{9}$$

(e) Suppose now we need to quantize both A_1 and A_2, and our bit budget is 2 bits. One way of doing this would be to use a one-bit Lloyd-Max quantizer derived in Part (b) for A_1, and use the same quantizer for A_2:

$$\hat{A}_2 = \begin{cases} q_1, & \text{if } A_2 < x_0 \\ q_2, & \text{if } A_2 \geq x_0, \end{cases}$$

where x_0, q_1, and q_2 are the numbers found in Part (b).

Consider an alternative procedure for quantizing the same variables, using the same number of bits:

$$\hat{A}_1' = \hat{A}_2' = \begin{cases} 1/3, & \text{if } A_1 + A_2 < 1 \\ 2/3, & \text{if } 1 \leq A_1 + A_2 < 2 \\ 4/3, & \text{if } 2 \leq A_1 + A_2 < 3 \\ 5/3, & \text{if } A_1 + A_2 \geq 3 \end{cases} \tag{10}$$

Such a quantization strategy where two or more variables are quantized jointly is called *vector quantization*, as opposed to *scalar quantization* which quantizes one variable at a time. The particular quantizer of Eq. (10) can be viewed as using both bits to quantize A_1 (since there are four quantization levels for A_1), and using zero bits to quantize A_2: once $\hat{A}_1'$ is stored, $\hat{A}_2'$ does not need to be stored at all, since it is equal to $\hat{A}_1'$.

For this new quantizer, find the mean-square errors,

$$E[(\hat{A}_1' - A_1)^2] \text{ and } E[(\hat{A}_2' - A_2)^2]. \tag{11}$$

Compare your results with Eq. (8) and (9), and explain why this new quantizer is able to achieve smaller errors with the same number of bits.

(**Hint**. For this problem, it may be very helpful to draw, in the a_1-a_2 plane, the set of all points where the joint PDF of A_1 and A_2 is non-zero. In the same picture, draw all possible points $(\hat{A}_1, \hat{A}_2)$. Then draw all possible points $(\hat{A}_1', \hat{A}_2')$.)

Solution. The joint PDF of A_1 and A_2 is illustrated in Fig. 9(a). It shows that the the random vector (A_1, A_2) can only appear in the shaded region, with uniform probability.

(a)

$$\begin{aligned} f_{A_1}(a_1) &= \int_{-\infty}^{\infty} f_{A_1,A_2}(a_1,a_2)da_2 \\ &= \begin{cases} 0 & a_1 < 0 \text{ or } a_1 > 2 \\ \int_0^1 \frac{1}{2}da_2 = \frac{1}{2} & 0 \leq a_1 \leq 1 \\ \int_1^2 \frac{1}{2}da_2 = \frac{1}{2} & 1 \leq a_1 \leq 2 \end{cases} \\ &= \frac{1}{2}[u(a_1) - u(a_1 - 2)]. \end{aligned}$$

This is a uniform distribution over $[0, 2]$.

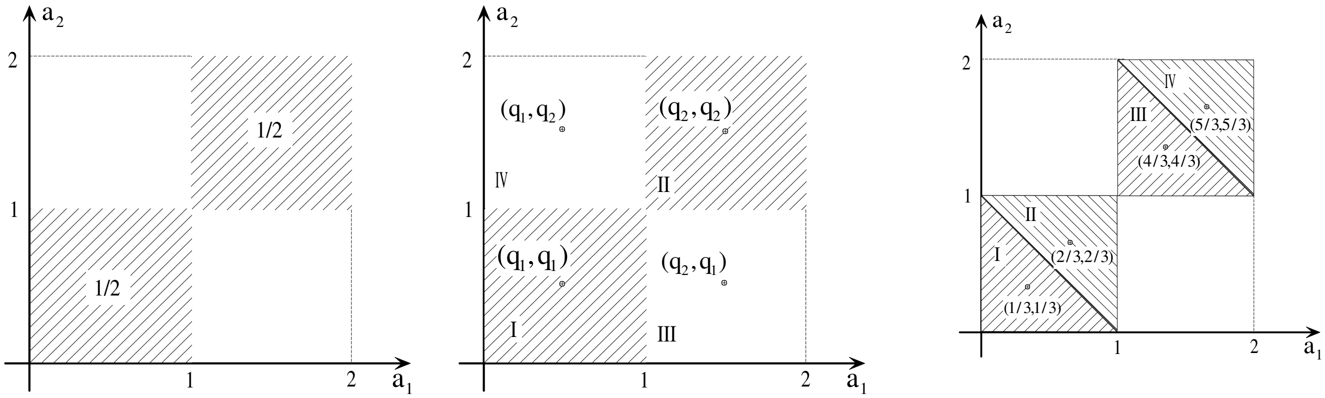


Figure 9: Problem 14.

(b) Since the marginal PDF of A_1 is uniform between 0 and 2, having $x_0 \leq 0$ would waste half of our bit on values that have probability zero. Having $x_0 \geq 2$ would be similarly wasteful. Hence $0 < x_0 < 2$. If $q_1 < 0$, the mean-square error can always be reduced by setting $q_1 = 0$. If $q_1 > x_0$, the mean-square error can be reduced by setting $q_1 = x_0$. Hence, $0 \leq q_1 \leq x_0$. A similar argument shows that $x_0 \leq q_2 \leq 2$. Therefore, the mean-square error can be rewritten as:

$$\begin{aligned}
E[(\hat{A}_1 - A_1)^2] &= \int_{-\infty}^{\infty} (\hat{a}_1 - a_1)^2 f_{A_1}(a_1) da_1 \\
&= \int_0^{x_0} (q_1 - a_1)^2 f_{A_1}(a_1) da_1 + \int_{x_0}^2 (q_2 - a_1)^2 f_{A_1}(a_1) da_1 \\
&= \frac{1}{2} \int_0^{x_0} (q_1 - a_1)^2 da_1 + \frac{1}{2} \int_{x_0}^2 (q_2 - a_1)^2 da_1. \qquad (12)
\end{aligned}$$

In order to minimize Eq. (12) over all possible x_0, q_1, and q_2, we take the partial derivative of (12) with respect to each of these variables and equate it to zero:

$$\begin{aligned}
\frac{\partial E[(\hat{A}_1 - A_1)^2]}{\partial x_0} &= \frac{1}{2}(q_1 - x_0)^2 - \frac{1}{2}(q_2 - x_0)^2 = 0, \\
\frac{\partial E[(\hat{A}_1 - A_1)^2]}{\partial q_1} &= \frac{1}{2} \int_0^{x_0} 2(q_1 - a_1) da_1 = 0, \\
\frac{\partial E[(\hat{A}_1 - A_1)^2]}{\partial q_2} &= \frac{1}{2} \int_{x_0}^2 2(q_2 - a_1) da_1 = 0.
\end{aligned}$$

After some simplifications, we get:

$$\begin{aligned}
x_0 &= (q_1 + q_2)/2, \qquad (13) \\
q_1 &= x_0/2, \qquad (14) \\
q_2 &= (x_0 + 2)/2. \qquad (15)
\end{aligned}$$

The solution to Eqs. (13)-(15) is:

$$x_0 = 1,\ q_1 = 0.5,\ q_2 = 1.5.$$

Evaluating the second derivatives of the MSE at this solution,

$$\begin{aligned}
\frac{\partial^2 E[(\hat{A}_1 - A_1)^2]}{\partial x_0^2} &= (x_0 - q_1) - (x_0 - q_2) = q_2 - q_1 = 1 > 0, \\
\frac{\partial^2 E[(\hat{A}_1 - A_1)^2]}{\partial q_1^2} &= \int_0^{x_0} da_1 = x_0 = 1 > 0, \\
\frac{\partial^2 E[(\hat{A}_1 - A_1)^2]}{\partial q_2^2} &= \int_{x_0}^{2} da_1 = 2 - x_0 = 1 > 0,
\end{aligned}$$

we see that the extremum we found is a minimum.

(c) With the values we obtained in Part (b), the mean-square error of the quantizer is:

$$\begin{aligned}
E[(\hat{A}_1 - A_1)^2] &= \frac{1}{2}\int_0^1 \left(\frac{1}{2} - a_1\right)^2 da_1 + \frac{1}{2}\int_1^2 \left(\frac{3}{2} - a_1\right)^2 da_1 \\
&= \frac{1}{12}.
\end{aligned}$$

(d) If a_1 and a_2 are interchanged, the joint PDF of A_1 and A_2 will remain the same. This symmetry means that the marginal PDF of A_2 is the same as the marginal PDF of A_1, and that the procedure to obtain the optimal quantizer for A_2 is also identical to part (b). Therefore, A_2 must have the same minimum MSE quantizer as A_1, with the same value for MSE.

(e) The new quantizer is illustrated in Fig. 9(c). The outcomes of the random vector in triangle I get quantized to $(1/3, 1/3)$, outcomes in triangle II get quantized to $(2/3, 2/3)$, etc. The MSE of quantizing A_1 can be broken into four integrals over the four triangles. Since the triangles are congruent, and since the positions of quantization points within the triangles are identical, the four integrals will be the same. Therefore, to get the MSE, we need to evaluate just one of the integrals and multiply the result by 4:

$$\begin{aligned}
E[(\hat{A}'_1 - A_1)^2] &= 4\int_I \left(\frac{1}{3} - a_1\right)^2 f_{A_1,A_2}(a_1, a_2)\,da_1 da_2 \\
&= 4\int_0^1 \int_0^{1-a_1} \left(\frac{1}{3} - a_1\right)^2 \frac{1}{2} da_2 da_1 \\
&= 2\int_0^1 \left(\frac{1}{3} - a_1\right)^2 (1 - a_1) da_1 \\
&= 2\int_0^1 \left(\frac{1}{3} - a_1\right)^3 da_1 + 2\int_0^1 \left(\frac{1}{3} - a_1\right)^2 \frac{2}{3} da_1 \\
&= -2\int_0^1 \left(a_1 - \frac{1}{3}\right)^3 da_1 + \frac{4}{3}\int_0^1 \left(a_1 - \frac{1}{3}\right)^2 da_1
\end{aligned}$$

$$
\begin{aligned}
&= -2\,\frac{\left(a_1-\frac{1}{3}\right)^4}{4}\Bigg|_0^1+\frac{4}{3}\,\frac{\left(a_1-\frac{1}{3}\right)^3}{3}\Bigg|_0^1 \\
&= -\frac{1}{2}\left(\frac{2}{3}\right)^4+\frac{1}{2}\left(-\frac{1}{3}\right)^4+\frac{4}{9}\left(\frac{2}{3}\right)^3-\frac{4}{9}\left(-\frac{1}{3}\right)^3 \\
&= -\frac{16}{2\cdot 81}+\frac{1}{2\cdot 81}+\frac{32}{3\cdot 81}+\frac{4}{3\cdot 81}=-\frac{15}{2\cdot 81}+\frac{36}{3\cdot 81}=\frac{24-15}{2\cdot 81} \\
&= \frac{1}{18}.
\end{aligned}
$$

By symmetry, $E[(\hat{A}_2'-A_2)^2]=E[(\hat{A}_1'-A_1)^2]=\frac{1}{18}$.

The quantizer of Parts (b)-(d) was designed separately for A_1 and A_2, based on the knowledge of the marginal PDF's of A_1 and A_2, but without exploiting the fact that A_1 and A_2 are dependent. This way, out of the four quantization points, $(1/2,1/2)$, $(1/2,3/2)$, $(3/2,1/2)$, and $(3/2,3/2)$, only two–$(1/2,1/2)$ and $(3/2,3/2)$–have a nonzero probability (see Fig. 9(b)). The other two points are wasted on coding impossible events. Therefore, our algorithm does not take full advantage of the two bits. The quantizer in Part (e) exploits dependence of A_1 and A_2, by trying to distribute the four quantization points evenly (more or less) throughout the region where the joint PDF of A_1 and A_2 is non-zero. Through this better placement of quantization points, it is able to achieve a smaller MSE.

Problem 15. *(Montier, [4].)*
You will represent Company A (the potential acquirer), which is currently considering acquiring Company T (the target) by means of a tender offer. The main complication is this: the value of Company T depends directly on the outcome of a major oil exploration project that is currently being undertaken. If the project fails, the company under current management will be worth nothing (\$0). But if the project succeeds, the value of the company under current management could be as high as \$100 per share. All share values between \$0 and \$100 are considered equally likely.

By all estimates, Company T will be worth considerably more in the hands of Company A than under current management. In fact, the company will be worth 50% more under the management of A than under the management of Company T. If the project fails, the company will be worth zero under either management. If the exploration generates a \$50 per share value, the value under Company A will be \$75. Similarly, a \$100 per share under Company T implies a \$150 value under Company A, and so on.

It should be noted that the only possible option to be considered is paying in cash for acquiring 100% of Company T's shares. This means that if you acquire Company T, its current management will no longer have any shares in the company, and therefore will not benefit in any way from the increase in its value under your management.

The board of directors of Company A has asked you to determine whether or not to submit an offer for acquiring company T's shares, and if so, what price they should offer for these shares. This offer must be made now, before the outcome of the drilling project is known. Company T will accept any offer from Company A, provided it is at a profitable price for them. It is also clear that Company T will delay its decision to accept or reject your bid until the results of the drilling project are in. Thus you (Company A) will not know the results of the exploration project when submitting your price offer, but Company T will know the results when deciding on your offer. As already explained, Company T is expected to accept any offer by Company A that is greater than the (per share) value of the company under current management, and to reject any offers that are below or equal to this value. Thus, if you offer \$60 per share, for example, Company T will accept if the value under current management is anything less than \$60. You are now requested to give advice to the representative of Company A who is deliberating over whether or not to submit an offer for acquiring Company T's shares, and if so, what price he/she should offer for these shares. If your advice is that he/she should not acquire Company T's shares, advise him/her to offer \$0 per share. If you think he/she should try to acquire Company T's shares, advise him/her to offer anything above \$0 per share. What is the offer that he/she should make? In other words, what is the optimal offer?

Solution. *(Ilya Pollak.)*
Let X be the share price after the drilling project results are known. Let y be our offer, in dollars. Note that $0 \leq y \leq 100$. This is because making an offer above \$100 does not make sense: Company T is never worth more than \$100 per share under the current management, and so an offer of \$100 or above is sure to be accepted. To compute Company A's expected profit per share, we consider two cases separately:

- Case 1: $y > X$. The offer is accepted, and Company A makes $1.5X - y$.
- Case 2: $y \leq X$. The offer is rejected, and Company A makes 0.

Based on the problem statement, to us X is a random variable. Using the total expectation theorem and the linearity of expectation, we have that the expected profit for Company A is:

$$\begin{aligned} E[1.5X - y|y > X] \cdot \mathbf{P}(y > X) + E[0|y \leq X] \cdot \mathbf{P}(y \leq X) &= (1.5E[X|X < y] - E[y|X < y])\mathbf{P}(X < y) \\ &= (1.5E[X|X < y] - y)\mathbf{P}(X < y) \end{aligned}$$

To evaluate this expression, we need to interpret the statement from the problem formulation that says: "All share values between \$0 and \$100 are considered equally likely." This implies a uniform distribution for the random variable X; however, there are several different reasonable assumptions that we can make. The random variable X can be modeled as either a continuous or a discrete uniform random variable. In the case of a discrete uniform distribution, we need to make further modeling assumptions regarding the granularity.

First, assuming that X is a continuous uniform random variable, we get that the conditional PDF of X given that $X < y$ is uniform between 0 and y, and therefore the conditional expectation of X given that $X < y$ is $y/2$. Hence, the expected profit is:

$$(1.5(y/2) - y)\mathbf{P}(X < y) = -0.25y\mathbf{P}(X < y).$$

Therefore, the largest possible profit is zero, and is achieved when $y = 0$. The optimal strategy is not to make an offer.

Alternatively, we can assume that prices change in increments of one cent, and therefore X is a discrete uniform random variable, with possible values being all integer multiples of \$0.01, from zero to \$100. In this case, the conditional PMF of X given that $X < y$, for any $y > 0$, is uniform between 0 and

$y - 0.01$, and therefore the conditional expectation of X given that $X < y$ is $y/2 - 0.005$. Hence, for any positive offer price y, the expected profit is:

$$(1.5(y/2 - 0.005) - y)\mathbf{P}(X < y) = -(0.25y + 0.0075)\mathbf{P}(X < y).$$

This is still negative, hence still the best strategy is not to make an offer.

Another reasonable alternative is to assume that prices change in increments of one dollar. In this case, X is a discrete uniform random variable, with possible values being all integers between 0 and \$100. The conditional PMF of X given $X < y$, for any $y > 0$, is uniform between 0 and $y - 1$, and therefore the conditional expectation of X given that $X < y$ is $y/2 - 0.5$. Hence, for any positive offer price y, the expected profit is:

$$(1.5(y/2 - 0.5) - y)\mathbf{P}(X < y) = -(0.25y + 0.75)\mathbf{P}(X < y).$$

This is again negative, hence still the best strategy is not to make an offer.

Problem 16. OPTIMAL PORTFOLIOS. *(Ilya Pollak.)*
This problem illustrates why it is good to diversify your investments. If you are able to split your investment among several independent assets, you will have a lower risk than if you invest everything in a single asset.

(a) You have \$1 to invest in two assets. Denote the yearly return from Asset 1 by R_1 and the yearly return from Asset 2 by R_2. This means that every dollar invested in Asset 1 will turn into $\$(1 + R_1)$ by the end of one year, and every dollar invested in Asset 2 will turn into $\$(1 + R_2)$. Suppose that R_1 and R_2 are independent continuous random variables with

$$\begin{aligned} E[R_1] = E[R_2] &= 0.1 \\ \mathrm{var}(R_1) &= 0.04 \\ \mathrm{var}(R_2) &= 0.09 \end{aligned}$$

How should you allocate your \$1 between the two assets in order to minimize risk, as measured by the standard deviation of your return?

Find the expected value and the standard deviation of the resulting return, and compare to the means and standard deviations for investing the entire \$1 into Asset 1 and into Asset 2.

Hint. Denote the amounts invested in Assets 1 and 2 by a and $1 - a$, respectively ($0 \le a \le 1$). Then the return of your portfolio is $R = aR_1 + (1 - a)R_2$. Find the value of a which minimizes the standard deviation of the portfolio return R. Note that minimizing the standard deviation is equivalent to minimizing the variance, but the expression for the variance is more convenient to differentiate. Then, for the value of a that you obtain, find $E[R]$ and $\mathrm{var}(R)$.

Solution. Using the independence of R_1 and R_2, we have:

$$\begin{aligned} \mathrm{var}(R) &= \mathrm{var}(aR_1 + (1 - a)R_2) \\ &= \mathrm{var}(aR_1) + \mathrm{var}((1 - a)R_2) \\ &= a^2\mathrm{var}(R_1) + (1 - a)^2\mathrm{var}(R_2) \\ &= 0.04a^2 + 0.09 - 0.18a + 0.09a^2 \\ &= 0.13a^2 - 0.18a + 0.09. \end{aligned}$$

This is a parabola and hence has a unique extremum. Since the coefficient of a^2 is positive, the extremum is a minimum. To find its location, take the derivative and set to zero:

$$\begin{aligned} 0.26a - 0.18 &= 0 \\ a &= \frac{9}{13} \approx 0.69 \end{aligned}$$

Thus we are investing $\$\frac{9}{13}$ into Asset 1, and $\$\frac{4}{13}$ into Asset 2, for a total return of

$$R = \frac{9}{13}R_1 + \frac{4}{13}R_2$$

Since expectations are linear, we have:

$$E[R] = \frac{9}{13}E[R_1] + \frac{4}{13}[R_2] = 0.1$$

Since the mean returns of Asset 1 and Asset 2 are 10% each, the mean portfolio return will always be 10%, regardless of how the two assets are weighted in the portfolio.

To find the variance, we substitute $a = 9/13$ into the formula derived above:

$$\begin{aligned} \mathrm{var}(R) &= a^2\mathrm{var}(R_1) + (1-a)^2\mathrm{var}(R_2) \\ &= \left(\frac{9}{13}\right)^2 \cdot 0.04 + \left(\frac{4}{13}\right)^2 \cdot 0.09 \\ &= \frac{3.24 + 1.44}{169} \\ &= \frac{468}{16900} \\ &= \frac{117}{4225} \approx 0.0277 \end{aligned}$$

Therefore, the standard deviation of the portfolio return is

$$\sigma_R = \sqrt{\frac{117}{4225}} \approx 0.166$$

Note that the standard deviations of Assets 1 and 2 are 0.2 and 0.3, respectively. Our portfolio has standard deviation of 0.166, which is lower than both the standard deviation of Asset 1 and the standard deviation of Asset 2. By mixing together two risky but independent assets, we have obtained a portfolio which is less risky than either one of the two assets, but has the same expected return.

(b) Assume the same means and variances for R_1 and R_2 as in Part (a), but now assume that R_1 and R_2 are correlated, with covariance

$$\mathrm{cov}(R_1, R_2) = 0.03,$$

and correlation coefficient

$$\rho_{R_1,R_2} = \frac{\mathrm{cov}(R_1, R_2)}{\sigma_{R_1}\sigma_{R_2}} = \frac{0.03}{0.2 \cdot 0.3} = 0.5.$$

Repeat Part (a) for these new conditions, i.e., find the portfolio weights a and $1-a$ that minimize the variance of the portfolio return, and find the resulting variance and standard deviation of the portfolio return.

Solution. Since $R = aR_1 + (1-a)R_2$, we have:

$$\begin{aligned} \text{var}(R) &= \text{var}(aR_1 + (1-a)R_2) \\ &= a^2\text{var}(R_1) + (1-a)^2\text{var}(R_2) + 2a(1-a)\text{cov}(R_1, R_2) \end{aligned} \tag{16}$$

Therefore,

$$\begin{aligned} \frac{d}{da}\text{var}(R) &= 2a\text{var}(R_1) - 2(1-a)\text{var}(R_2) + (2-4a)\text{cov}(R_1, R_2) \\ &= 2[a(\text{var}(R_1) + \text{var}(R_2) - 2\text{cov}(R_1, R_2)) - \text{var}(R_2) + \text{cov}(R_1, R_2)] \end{aligned}$$

Setting the derivative to zero yields:

$$\begin{aligned} a(\text{var}(R_1) + \text{var}(R_2) - 2\text{cov}(R_1, R_2)) &= \text{var}(R_2) - \text{cov}(R_1, R_2) \\ a &= \frac{\text{var}(R_2) - \text{cov}(R_1, R_2)}{\text{var}(R_1) + \text{var}(R_2) - 2\text{cov}(R_1, R_2)} \\ &= \frac{0.09 - 0.03}{0.04 + 0.09 - 2 \cdot 0.03} \\ &= \frac{0.06}{0.07} = \frac{6}{7} \end{aligned}$$

The second derivative is:

$$\begin{aligned} \frac{d^2}{da^2}\text{var}(R) &= 2(\text{var}(R_1) + \text{var}(R_2) - 2\text{cov}(R_1, R_2)) \\ &= 2(0.04 + 0.09 - 2 \cdot 0.03) = 0.14 > 0, \end{aligned}$$

and therefore the extremum we found is a minimum. Substituting $a = 6/7$ into Eq. (16), we get:

$$\begin{aligned} \text{var}(R) &= \left(\frac{6}{7}\right)^2 \cdot 0.04 + \left(\frac{1}{7}\right)^2 \cdot 0.09 + 2 \cdot \frac{6}{7} \cdot \frac{1}{7} \cdot 0.03 \\ &= 0.01 \cdot \frac{144 + 9 + 36}{49} = \frac{189}{4900} \approx 0.0386. \end{aligned}$$

Therefore, the standard deviation of the portfolio return is now

$$\sigma_R = \sqrt{\frac{189}{4900}} \approx 0.196.$$

This is still a little smaller than the standard deviation 0.2 of Asset 1; however, since Asset 2 is now significantly correlated with Asset 1 (correlation coefficient 0.5), it does not provide much diversification to our portfolio.

(c) Suppose three assets have returns R_1, R_2, and R_3 which are independent random variables with means $E[R_1] = 0.05$, $E[R_2] = 0.1$, and $E[R_3] = 0.15$, and variances $\text{var}(R_1) = 0.01$,

$\text{var}(R_2) = 0.04$, and $\text{var}(R_3) = 0.09$. Suppose you want to invest \$1 into these three assets so that your expected return is 0.1, and your risk is as small as possible. In other words, you invest a_1 in Asset 1, $1 - a_1 - a_3$ in Asset 2, and a_3 in Asset 3, and obtain a portfolio whose return is $R = a_1 R_1 + (1 - a_1 - a_3)R_2 + a_3 R_3$. Find a_1 and a_3 to minimize $\text{var}(R)$ subject to the constraint $E[R] = 0.1$. For these values of a_1 and a_3, find the variance and standard deviation of the portfolio return R.

Solution. The expected portfolio return is:

$$\begin{aligned} E[R] &= a_1 E[R_1] + (1 - a_1 - a_3)E[R_2] + a_3 E[R_3] \\ &= 0.05a_1 + 0.1(1 - a_1 - a_3) + 0.15a_3 \\ &= 0.1 - 0.05a_1 + 0.05a_3. \end{aligned}$$

In order for the expected portfolio return to be equal to 0.1, we must therefore have $a_1 = a_3$. If this is the case, we have that the variance of the portfolio return is:

$$\begin{aligned} \text{var}(R) &= a_1^2 \text{var}(R_1) + (1 - 2a_2)^2 \text{var}(R_2) + a_1^2 \text{var}(R_3) \\ &= a_1^2[\text{var}(R_1) + 4\text{var}(R_2) + \text{var}(R_3)] - 4a_1 \text{var}(R_2) + \text{var}(R_2) \\ &= a_1^2[0.01 + 0.16 + 0.09] - 0.16a_1 + 0.04 \\ &= 0.26a_1^2 - 0.16a_1 + 0.04. \end{aligned}$$

Differentiating and setting the derivative to zero yields $0.52a_1 - 0.16 = 0$, and therefore the solution is $a_1 = 0.16/0.52 = 4/13$. Since the second derivative is positive, this is a minimum. The weights for Assets 3 and 2 are $a_3 = a_1 = 4/13$ and $a_2 = 1 - a_1 - a_3 = 5/13$. For these weights, the variance of the portfolio return is:

$$\begin{aligned} \text{var}(R) &= 0.26\left(\frac{4}{13}\right)^2 - 0.16 \cdot \frac{4}{13} + 0.04 \\ &= 0.02 \cdot \frac{16}{13} - \frac{0.64}{13} + \frac{0.52}{13} \\ &= \frac{0.2}{13} = \frac{1}{65} \approx 0.0154. \end{aligned}$$

The standard deviation is

$$\sigma_R = \sqrt{\frac{1}{65}} \approx 0.124.$$

(d) Suppose you are now allocating your \$1 among ten assets with independent returns. The expected return for each asset is 0.1. The return variances are $1, 1/2, 1/3, \ldots, 1/10$. Find the allocation weights a_1, a_2, $\ldots$, a_9, $1 - a_1 - a_2 - \ldots - a_9$ that minimize the variance of your portfolio return. Find the variance and standard deviation of the portfolio return for these weights.

Solution. Since the returns of individual assets are independent, we have:

$$\begin{aligned} \text{var}(R) &= \sum_{i=1}^{10} a_i^2 \text{var}(R_i) \\ &= \sum_{i=1}^{9} a_i^2 \text{var}(R_i) + \left(1 - \sum_{i=1}^{9} a_i\right)^2 \text{var}(R_{10}), \end{aligned}$$

where we used the notation

$$a_{10} = 1 - \sum_{i=1}^{9} a_i.$$

To get the minimum of the variance, we set to zero its partial derivatives with respect to $a_1, \ldots, a_9$ and then check that the second partial derivatives are positive. The first partial derivative with respect to a_i is:

$$\begin{aligned} \frac{\partial}{\partial a_i}\mathrm{var}(R) &= 2a_i\mathrm{var}(R_i) + 2\left(1 - \sum_{i=1}^{9} a_i\right)(-1)\mathrm{var}(R_{10}) \\ &= 2a_i\mathrm{var}(R_i) - 2a_{10}\mathrm{var}(R_{10}). \end{aligned}$$

For this to be zero for all $i = 1, \ldots, 9$, we therefore must have:

$$a_i = a_{10}\frac{\mathrm{var}(R_{10})}{\mathrm{var}(R_i)}, \tag{17}$$

for $i = 1, \ldots, 9$. Note that this identity also holds for $i = 10$. Since the sum of the ten weights must be equal to one, we obtain the following by summing the above equation over $i = 1, \ldots, 10$:

$$1 = a_{10}\mathrm{var}(R_{10}) \sum_{i=1}^{10} \frac{1}{\mathrm{var}(R_i)},$$

and hence

$$a_{10} = \frac{1}{\mathrm{var}(R_{10}) \sum_{i=1}^{10} \frac{1}{\mathrm{var}(R_i)}}.$$

Substituting this into Eq. (17), we get:

$$\begin{aligned} a_i &= \frac{1}{\mathrm{var}(R_i) \sum_{j=1}^{10} \frac{1}{\mathrm{var}(R_j)}} \\ &= \frac{1}{\frac{1}{i} \sum_{j=1}^{10} j} = \frac{i}{55}. \end{aligned}$$

The second partial derivative of the variance is

$$\frac{\partial^2}{\partial a_i^2}\mathrm{var}(R) = 2\mathrm{var}(R_i) + 2\mathrm{var}(R_{10}) > 0,$$

which means that we found a minimum. Substituting the values of the weights we found into the expression for the variance, we get the following value of the variance for our optimal portfolio:

$$\begin{aligned}
\text{var}(R) &= \sum_{i=1}^{10} \left(\frac{i}{55}\right)^2 \text{var}(R_i) \\
&= \sum_{i=1}^{10} \left(\frac{i}{55}\right)^2 \frac{1}{i} \\
&= \sum_{i=1}^{10} \frac{i}{55^2} \\
&= \frac{1}{55},
\end{aligned}$$

which is significantly lower than the variance of any single asset in the portfolio. The standard deviation is:

$$\sigma_R = \sqrt{\frac{1}{55}} \approx 0.135.$$

Problem 17. SIMULATING A FAIR COIN WITH A BIASED ONE. *(After Bertsekas-Tsitsiklis [1] and Stout-Warren [5].)*

An NFL referee has only one biased coin available for the pre-game coin toss. For this coin, the probability of heads is 0.6, and both the referee and the two teams know this. The referee would like to devise a coin-tossing procedure that each team will have a 0.5 probability of winning. Problem 1.33 in the recommended text [1] suggests the following strategy: Have one team call HT or TH and flip the coin twice. If the result is HH or TT, the procedure is repeated.

(a) What is the probability to have to keep going after the n-th independent repetition of this procedure? In other words, what is the probability that no single trial among the n independent trials results in HT or TH?

Solution. If a biased coin with probability of heads p is independently tossed twice, the probability of HH is p^2 and the probability of TT is $(1-p)^2$. The probability to keep getting HH or TT in n independent repetitions of this procedure is therefore $(p^2 + (1-p)^2)^n$. In our case, this is $(0.6^2 + 0.4^2)^n = 0.52^n$.

(b) What is the probability to never terminate, i.e., to keep getting HH or TT forever?

Solution. The probability to keep getting HH or TT forever is $\lim_{n\to\infty}(p^2 + (1-p)^2)^n$ which is zero, as long as $0 < p < 1$.

(c) What is the expected number of repetitions of this procedure required to terminate, i.e., to get an outcome that's either HT or TH? What is the expected number of coin flips required to terminate this procedure?

Solution. The number of repetitions to achieve an HT or TH is a geometric random variable with probability of success $2p(1-p) = 0.48$ which means that its expectation is $1/0.48 \approx 2.08$.

Since each repetition of the procedure consists of two coin flips, the expected number of coin flips is $2 \cdot 2.08 = 4.16$.

(d) There exist many other more complicated algorithms that are more efficient—i.e., achieve a smaller expected value. These algorithms exploit the fact that there exist other equiprobable sequences, for example, HHTT and TTHH. (See, e.g., [5].) For instance, the first team could bet on HT, HHHT, TTHT or HHTT, and the second on TH, HHTH, TTTH or TTHH. If none of these come up, the procedure is repeated. This is still fair: each team has probability $p(1-p)+p^2(1-p)^2+p^3(1-p)+p(1-p)^3$ of winning in one round of the procedure. Calculate the expected number of flips for this strategy. One way of doing this it to follow these steps:

(i) Find the probabilities R_2 and R_4 to terminate after two and four flips, respectively.
Solution.

$$\begin{aligned} R_2 &= 2p(1-p) = 0.48 \\ R_4 &= 2(p^2(1-p)^2 + p^3(1-p) + p(1-p)^3) = 2(0.0576 + 0.0864 + 0.0384) = 0.3648 \end{aligned}$$

(ii) Let $R_T = R_2 + R_4$ be the probability to terminate in one round. Let N be the total number of rounds required to terminate (i.e., to determine the winner.) Express $E[N]$ as a function of R_T.
Solution. The total number of required rounds is geometric with parameter R_T, and therefore the expected total number of rounds is $1/R_T$.

(iii) Given that $N = n$, find the conditional probabilities that the N-th round consists of two flips and that the N-th round consists of four flips.
Solution. Given that we have to go exactly N rounds, the conditional probability that the N-th round consists of two flips is $R_2/(R_2 + R_4) \approx 0.568$, and the conditional probability that the N-th round consists of four flips is $R_4/(R_2 + R_4) \approx 0.432$.

(iv) If the total number of rounds is $N = n$, then the total number of flips is $4(n-1)$ plus the number of flips in the last round. Use this observation in conjunction with the result of part (iii) to find the conditional expectation of the total number of flips N_f given that the total number of rounds is $N = n$.
Solution. Using the total expectation theorem, the conditional expectation of the number of flips given $N = n$ rounds is:

$$E[N_f|N=n] = \frac{R_2}{R_2+R_4}(4n-2) + \frac{R_4}{R_2+R_4}(4n) = \frac{4n(R_2+R_4) - 2R_2}{R_2+R_4} \approx 4n - 1.136.$$

(v) Use the iterated expectation formula to find the expected number of flips $E[N_f]$.
Solution. Using the iterated expectation formula, the expected number of flips is:

$$\begin{aligned} E[E[N_f|N]] &= \frac{4E[N](R_2+R_4) - 2R_2}{R_2+R_4} \\ &= \frac{4 - 2R_2}{R_2+R_4} \\ &= \frac{4 - 2 \cdot 0.48}{0.8448} = \frac{3.04}{0.8448} \approx 3.60, \end{aligned}$$

which is about 13% faster than the 4.16 flips expected under the simpler strategy.

Problem 18. *(Ilya Pollak.)*
The CDF of a continuous random variable X is:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x/2, & 0 \le x \le 2 \\ 1, & x > 2 \end{cases}$$

a. Find the PDF of X.

   Solution. Differentiating the CDF, we get:

$$f_X(x) = \begin{cases} 0, & x < 0 \\ 1/2, & 0 < x < 2 \\ 0, & x > 2 \end{cases}$$

   Thus, X is a continuous random variable, uniformly distributed between 0 and 2. Note that the derivative of the CDF does not exist at $x = 0$ and at $x = 2$.

b. Find $E[X]$.

   Solution. $E[X] = (0 + 2)/2 = 1$.

c. Find $\mathrm{var}(X)$.

   Solution. $\mathrm{var}(X) = (2 - 0)^2/12 = 1/3$.

d. Find $E[X^3]$.

   Solution.

$$E[X^3] = \int_0^2 x^3/2 \, dx = \left. \frac{x^4}{8} \right|_0^2 = 16/8 = 2.$$

e. Find the conditional PDF of X given that $0 < X < 1/2$.

   Solution. Given that $0 < X < 1/2$, the conditional PDF of X is zero outside of the region $0 < x < 1/2$. Inside the region, it is equal to the PDF of X divided by the probability of $0 < X < 1/2$. For $0 < x < 1/2$, the PDF is $1/2$ as found in Part a. $\mathbf{P}(0 < X < 1/2) = \int_0^{1/2} 1/2 \, dx = 1/4$. Hence, we get:

$$f_{X|0<X<1/2}(x) = \begin{cases} 0, & x < 0 \\ 2, & 0 < x < 1/2 \\ 0, & x > 1/2 \end{cases}$$

   An alternative way of arriving at the same answer is simply to notice that, since the PDF of X is uniform over $0 \le x \le 2$, the conditional PDF of X given $0 < X < 1/2$ must be uniform over $0 < x < 1/2$.

f. Find the PDF of the random variable $Y = X - 1$.

   Solution. First, find the CDF:

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(X - 1 \le y) = \mathbf{P}(X \le y + 1) = F_X(y + 1)$$

Then differentiate to find the PDF:

$$f_Y(y) = \frac{d}{dy}F_X(y+1) = f_X(y+1) = \begin{cases} 0, & y < -1 \\ 1/2, & -1 < y < 1 \\ 0, & y > 1 \end{cases}$$

As a quick check, note that, since X is between 0 and 2 with probability 1, $Y = X - 1$ must be between -1 and 1 with probability 1.

An alternative method is to note that by subtracting 1 from a random variable, we move its PDF to the left by 1. Hence, the PDF of Y is uniform from -1 to 1.

Problem 19. *(Ilya Pollak.)*
Consider the following joint PDF of two continuous random variables, X and Y:

$$f_{XY}(x,y) = \begin{cases} 1, & 0 \le x \le 1 \quad \text{and} \quad 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

a. Find the marginal PDF $f_X(x)$.

Solution. To get the marginal density of X, integrate the joint density:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy \\ &= \int_{-\infty}^{\infty} (u(x) - u(x-1))(u(y) - u(y-1))dy \\ &= (u(x) - u(x-1)) \int_{-\infty}^{\infty} (u(y) - u(y-1))dy \\ &= (u(x) - u(x-1)) \int_{0}^{1} 1\, dy \\ &= u(x) - u(x-1). \end{aligned}$$

So X is uniformly distributed on [0,1].

b. Specify all values y for which the conditional PDF $f_{X|Y}(x|y)$ exists. For all these values, find $f_{X|Y}(x|y)$.

Solution. The conditional PDF $f_{X|Y}(x|y)$ exists for all values of y for which $f_Y(y) > 0$. The marginal distribution of Y is uniform over [0,1]. This can be verified either by integrating the joint density over x, or by noticing that the joint density is symmetric (in the sense that swapping X and Y does not change the joint density) and therefore Y must have exactly the same marginal density as the marginal density of X, which was found in Part a. Thus, $f_{X|Y}(x|y)$ exists for $0 \le y \le 1$. For these values of y,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = u(x) - u(x-1),$$

i.e., the conditional distribution of X given any value $Y = y \in [0,1]$ is uniform between 0 and 1.

c. Are X and Y independent?

Solution. Since $f_X(x)f_Y(y) = f_{XY}(x, y)$, X and Y are independent.

d. Find the probability $\mathbf{P}(X > 3/4)$.

Solution. This probability is the integral from $3/4$ to ∞ of the uniform density found in Part a, which is $1/4$.

e. Find the probability $\mathbf{P}(X < Y)$.

Solution. Since the joint density of X and Y is symmetric, $\mathbf{P}(X < Y) = \mathbf{P}(X > Y)$. But $\mathbf{P}(X < Y) + \mathbf{P}(X > Y) = 1$, and hence $\mathbf{P}(X < Y) = 1/2$.

f. Find the conditional probability $\mathbf{P}(X > 3/4|Y < 1/2)$.

Solution. Since the two random variables are independent, the conditional probability of $X > 3/4$ given $Y < 1/2$ is the same as the unconditional probability of $X > 3/4$, which was found to be $1/4$ in Part d.

Problem 20. Y is a random variable with mean 0 and variance 2. Z is a random variable with mean 0 and variance 18. The correlation coefficient of Y and Z is $\rho(Y, Z) = 0.5$.

a. Find $E[Y + Z]$.

Solution. $E[Y + Z] = E[Y] + E[Z] = 0$.

b. Find $E[Z^2 + 1]$.

Solution. $E[Z^2 + 1] = E[Z^2] + 1 = \text{var}(Z) + (E[Z])^2 + 1 = 18 + 0 + 1 = 19$.

c. Find $E[YZ]$.

Solution. $\text{cov}(Y, Z) = \rho(Y, Z)\sigma_Y\sigma_Z = 0.5 \cdot \sqrt{2} \cdot \sqrt{18} = 3$. Therefore,

$$E[YZ] = \text{cov}(Y, Z) + E[Y]E[Z] = \text{cov}(Y, Z) = 3.$$

d. Find $\text{var}(Y + Z)$.

Solution. $\text{var}(Y + Z) = \text{var}(Y) + \text{var}(Z) + 2\text{cov}(Y, Z) = 2 + 18 + 6 = 26$.

Problem 21. X is a normal (Gaussian) random variable with mean 0.

a. Find the probability of the event $\{X \text{ is an integer}\}$.

Solution. $\mathbf{P}(X = n) = \int_n^n f_X(x)\,dx = 0$ for any n, and therefore

$$\mathbf{P}(X \text{ is an integer}) = \sum_{n=-\infty}^{\infty} \mathbf{P}(X = n) = 0.$$

b. List the following probabilities in the increasing order, from smallest to largest: $\mathbf{P}(-1 \le X \le 1)$, $\mathbf{P}(0 \le X \le 1)$, $\mathbf{P}(0 \le X \le 2)$, $\mathbf{P}(1 \le X \le 2)$.

Solution. Since a normal PDF with zero mean is an even function,

$$\mathbf{P}(-1 \le X \le 0) = \int_{-1}^{0} f_X(x)\,dx = \int_{-1}^{0} f_X(-x)\,dx = \int_{0}^{1} f_X(x)\,dx = \mathbf{P}(0 \le X \le 1). \tag{18}$$

Since a normal PDF with zero mean is monotonically decreasing for $x \ge 0$,

$$\mathbf{P}(0 \le X \le 1) > \mathbf{P}(1 \le X \le 2). \tag{19}$$

Since the event $\{0 \le X \le 2\}$ is the union of the events $\{0 \le X \le 1\}$ and $\{1 \le X \le 2\}$ each of which has nonzero probability, we have:

$$\mathbf{P}(0 \le X \le 1) < \mathbf{P}(0 \le X \le 2).$$

Finally,

$$\begin{aligned}
\mathbf{P}(0 \le X \le 2) &= \mathbf{P}(0 \le X \le 1) + \mathbf{P}(1 \le X \le 2) \\
&\overset{\text{Eq. (19)}}{<} 2\mathbf{P}(0 \le X \le 1) \\
&\overset{\text{Eq. (18)}}{=} \mathbf{P}(0 \le X \le 1) + \mathbf{P}(-1 \le X \le 0) \\
&= \mathbf{P}(-1 \le X \le 1).
\end{aligned}$$

Putting everything together, we have the following answer:

$$\mathbf{P}(1 \le X \le 2) < \mathbf{P}(0 \le X \le 1) < \mathbf{P}(0 \le X \le 2) < \mathbf{P}(-1 \le X \le 1)$$

Problem 22. Let c be a fixed point in a plane, and let $\mathcal{C}$ be a circle of length 1 centered at c. In other words, $\mathcal{C}$ is the set of all points in the plane whose distance from c is $1/(2\pi)$. (Note that $\mathcal{C}$ does *NOT* include its interior, i.e., the points whose distance from c is smaller than $1/(2\pi)$.) Points R_1 and R_2 are each uniformly distributed over the circle, and are independent. (The fact that R_1 and R_2 are uniformly distributed over a circle of unit length means that, for any arc $\mathcal{A}$ whose length is p, the following identities hold: $\mathbf{P}(R_1 \in \mathcal{A}) = p$ and $\mathbf{P}(R_2 \in \mathcal{A}) = p$.) Point r is a non-random, fixed point on the circle.

We let $\overset{\frown}{R_1R_2}$ be the arc with endpoints R_1 and R_2, obtained by traversing the circle clockwise from R_1 to R_2.

We let $\overset{\frown}{R_1rR_2}$ be the arc with endpoints R_1 and R_2 that contains the point r:

$$\overset{\frown}{R_1rR_2} = \begin{cases} \overset{\frown}{R_1R_2}, & \text{if } r \in \overset{\frown}{R_1R_2} \\ \overset{\frown}{R_2R_1}, & \text{if } r \notin \overset{\frown}{R_1R_2} \end{cases}$$

a. The random variable L is defined as the length of the arc $\overset{\frown}{R_1R_2}$. Find $E[L]$.

Solution. Let L_1 be the length of the arc $\overset{\frown}{R_2R_1}$. Due to symmetry between R_1 and R_2, $E[L_1] = E[L]$. But the sum of the lengths of the two arcs is equal to the length of the circle: $L + L_1 = 1$. Hence, we have $E[L] + E[L_1] = 1$, and $E[L] = 1/2$.

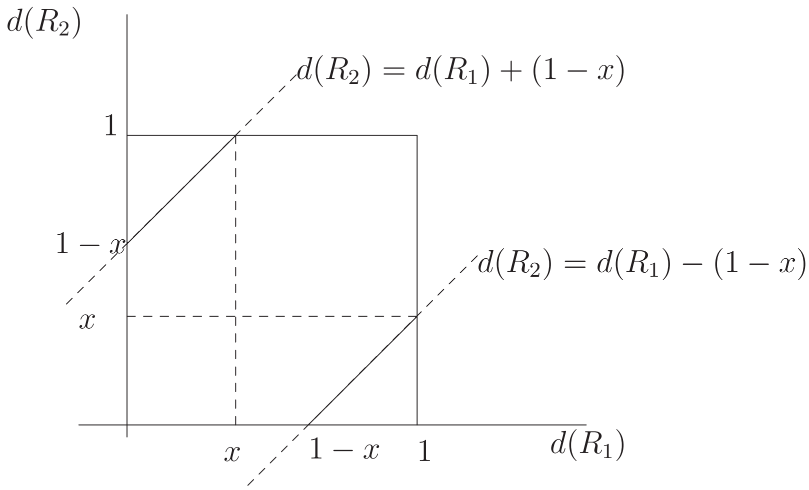


Figure 10:

b. The random variable L' is defined as the length of the arc $R_1\widehat{r}R_2$. Find $E[L']$.

Solution. For any point $x \in \mathcal{C}$, let $d(x)$ be the length of the arc $\widehat{rx}$, i.e., the distance between r and x along the circle in the clockwise direction. Since R_1 and R_2 are independent uniform points on the circle, $d(R_1)$ and $d(R_2)$ are independent random variables, each uniformly distributed between 0 and 1. Then the length of the arc between R_1 and R_2 that does *not* contain r, is $|d(R_1) - d(R_2)|$. The length of the arc that contains r is $L' = 1 - |d(R_1) - d(R_2)|$. Since both $d(R_1)$ and $d(R_2)$ are between 0 and 1 with probability 1, L' is also between 0 and 1 with probability 1. To get its probability density function, we first find its CDF on the interval $[0, 1]$, i.e., $\mathbf{P}(L' \leq x)$:

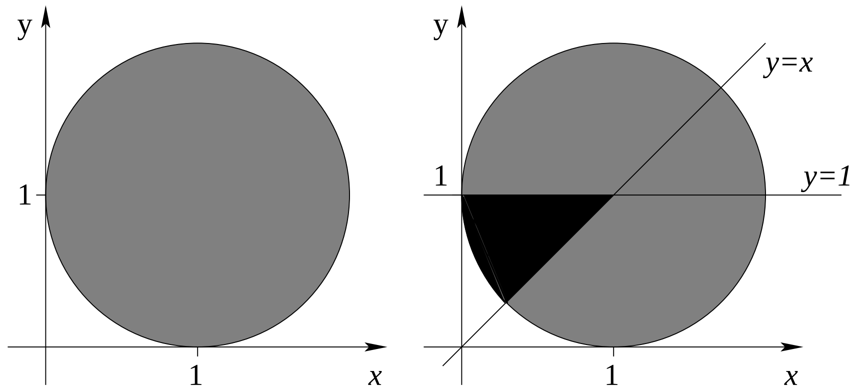
$$\begin{aligned}
\mathbf{P}(L' \leq x) &= \mathbf{P}(1 - |d(R_1) - d(R_2)| \leq x) \\
&= \mathbf{P}(1 - |d(R_1) - d(R_2)| \leq x) \\
&= \mathbf{P}(|d(R_1) - d(R_2)| \geq 1 - x) \\
&= \mathbf{P}(d(R_1) - d(R_2) \geq 1 - x) + \mathbf{P}(d(R_2) - d(R_1) \geq 1 - x)
\end{aligned}$$

The joint distribution of $d(R_1)$ and $d(R_2)$ is uniform over the square $[0, 1] \times [0, 1]$ shown in Fig. 10. Since the square has unit area, the probability of any set is the area of its intersection with the square. The event $d(R_1) - d(R_2) \geq 1 - x$ corresponds to the triangle in the lower-right corner of the square in Fig. 10. Therefore, its probability is the area of the triangle, which is $x^2/2$. Similarly, the probability of the event $d(R_2) - d(R_1) \geq 1 - x$ is the area of the upper-left triangle, which is also $x^2/2$. We therefore have:

$$\mathbf{P}(L' \leq x) = x^2.$$

Differentiating, we get the PDF for L' on the interval $[0, 1]$: $2x$. The expectation is therefore equal to:

$$E[L'] = \int_0^1 x \cdot 2x \, dx = \left. \frac{2x^3}{3} \right|_0^1 = \frac{2}{3}.$$



(a) The support of $f_{X,Y}$. (b) Illustration to Part d.

Figure 11: Illustrations for Problem 23.

It may seem counterintuitive that the answers to Parts a and b are different. The reason that $E[L'] > 1/2$ is that the arc connecting R_1 and R_2 and covering an arbitrary point r is likely to be larger than the arc connecting R_1 and R_2 that does not cover r.

Problem 23. *(Ilya Pollak.)*
Random variables X and Y have the following joint probability density function:

$$f_{X,Y}(x, y) = \begin{cases} C, & (x-1)^2 + (y-1)^2 \le 1, \\ 0, & \text{otherwise,} \end{cases}$$

where C is a positive real number. In other words, the point (X, Y) is uniformly distributed over the circle of radius 1, centered at (1,1).

a Find the conditional probability density $f_{X|Y}(x|1)$.

b Find the conditional expectation $E[X|Y = 0.3]$.

c Can the marginal probability density $f_X(x)$ be normal (Gaussian)? Explain without calculating $f_X(x)$.

d Find the conditional probability $\mathbf{P}(Y > X|Y \le 1)$.

e Find the PDF of random variable Z defined by: $Z = \sqrt{(X-1)^2 + (Y-1)^2}$.

Solution.

a. Given that $Y = 1$, X ranges from 0 to 2. Since the joint density is uniform, the conditional density of X is also uniform:

$$f_{X|Y}(x|1) = \begin{cases} 0.5, & 0 \le x \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

b. For any value of Y, the conditional density of X is uniform and symmetric about $X = 1$. Therefore $E[X|Y = 0.3] = 1$.

c. The marginal probability density of X must be identically zero outside of the interval [0,2]. Since a normal density is strictly greater than zero on the whole real line, f_X cannot be normal.

d. The conditioning event $Y \leq 1$ corresponds to all the points below the line $y = 1$ in Fig. 11(b), i.e., the bottom half of the circle. Among those points, the black region–which is one-eighth of the circle–corresponds to the event $Y > X$. Since the joint distribution is uniform, the conditional probability is just $(1/8)/(1/2) = 1/4$.

e. First note that, for $z < 0$, $f_Z(z) = 0$. For $z \geq 0$, the event that $Z \leq z$ means that the point (X, Y) lies in the circle of radius z, centered at (1,1). When $z > 1$, the probability of this event is 1, and therefore $f_Z(z) = 0$. When $0 \leq z \leq 1$, this is the ratio of the area of the circle with radius z and the area of the circle of radius 1:

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \mathbf{P}(Z \leq z) = \frac{d}{dz} \left(\frac{\pi z^2}{\pi \cdot 1^2} \right) = 2z.$$

Putting all three cases together, we get:

$$f_Z(z) = \begin{cases} 2z, & 0 \leq z \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(Note that $F_Z(z)$ is not differentiable at $z = 1$, and therefore, strictly speaking, the value of f_Z at $z = 1$ is undefined.)

Problem 24. *(Ilya Pollak.)*
Let Z be a continuous random variable, uniformly distributed between 0 and 2π. Are random variables $X = \sin Z$ and $Y = \cos Z$ uncorrelated? Are they independent? (Recall that X and Y are said to be uncorrelated if $E[(X - E[X])(Y - E[Y])] = 0$.)

Solution. We first find the expectations of X and Y and then show that cov(X,Y) = 0.

$$\begin{aligned} E[X] &= \int_0^{2\pi} \frac{1}{2\pi} \sin z \; dz = 0, \\ E[Y] &= \int_0^{2\pi} \frac{1}{2\pi} \cos z \; dz = 0, \\ \operatorname{cov}(X, Y) &= E[XY] = E[\sin Z \cos Z] = \frac{1}{2} E[\sin 2Z] = \frac{1}{2} \int_0^{2\pi} \frac{1}{2\pi} \sin 2z \; dz = 0, \end{aligned}$$

therefore, X and Y are uncorrelated. However, note that the knowledge of Y gives us information about X. For example, given that $Y = -1$, we know that Z must be equal to π, and therefore $X = 0$: $\mathbf{P}(X = 0|Y = -1) = 1$. However, given that $Y = 0$, Z is either $\pi/2$ or $3\pi/2$, and therefore $X \neq 0$: $\mathbf{P}(X = 0|Y = 0) = 0$. Thus, the conditional distribution for X given $Y = y$ depends on y, which means that X and Y are not independent.

Problem 25. *(Ilya Pollak.)*
Random variables B and C are jointly uniform over a $2\ell \times 2\ell$ square centered at the origin, i.e., B

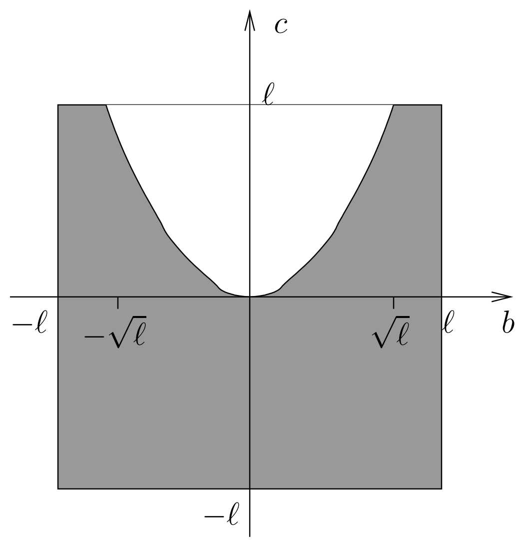


Figure 12: Illustration to Problem 25.

and C have the following joint probability density function:

$$f_{B,C}(b,c) = \begin{cases} \frac{1}{4\ell^2}, & -\ell \le b \le \ell \text{ and } -\ell \le c \le \ell, \\ 0, & \text{otherwise.} \end{cases}$$

It is given that $\ell \ge 1$. Find the probability that the quadratic equation $x^2 + 2Bx + C = 0$ has real roots. (Your answer will be an expression involving ℓ.) What is the limit of this probability as $\ell \to \infty$?

Solution. The quadratic equation has real roots if and only if $B^2 - C \ge 0$. To find the probability of this event, we integrate the joint density over all points (b, c) of the $2\ell \times 2\ell$ square for which $b^2 - c \ge 0$, i.e., over the gray set in Fig. 12. To do this, it is easier to integrate over the white portion of the square and subtract the result from 1:

$$\begin{aligned} 1 - \int_{-\sqrt{\ell}}^{\sqrt{\ell}} \int_{b^2}^{\ell} \frac{1}{4\ell^2}\, dc\, db &= 1 - \int_{-\sqrt{\ell}}^{\sqrt{\ell}} \frac{\ell - b^2}{4\ell^2}\, db = 1 - \left.\frac{b}{4\ell}\right|_{-\sqrt{\ell}}^{\sqrt{\ell}} + \left.\frac{b^3}{12\ell^2}\right|_{-\sqrt{\ell}}^{\sqrt{\ell}} \\ &= 1 - \frac{2\sqrt{\ell}}{4\ell} + \frac{2\ell\sqrt{\ell}}{12\ell^2} = 1 - \frac{1}{3\sqrt{\ell}}. \end{aligned}$$

As $\ell \to \infty$, this tends to 1.

Problem 26. *(Ilya Pollak.)*
A stick of unit length is broken into two at random, i.e., the location of the breakpoint is uniformly distributed between the two ends of the stick.

a What is the expected length of the smaller piece?

b What is the expected value of the ratio $\dfrac{\text{length of the smaller piece}}{\text{length of the larger piece}}$? (You can use $\ln 2 \approx 0.69$.)

Solution.

a Let the length of the left piece be X, and the length of the smaller piece be Y. For $X \leq 1/2$, $Y = X$, and for $X > 1/2$, $Y = 1 - X$. In other words, $f_{Y|X\leq 1/2}(y) = f_{X|X\leq 1/2}(y)$ which is uniform between 0 and 1/2, resulting in $E[Y|X \leq 1/2] = 1/4$, and $f_{Y|X>1/2}(y) = f_{1-X|X>1/2}(y)$ which is also uniform between 0 and 1/2, resulting in $E[Y|X > 1/2] = 1/4$. Therefore, using the total expectation theorem,

$$E[Y] = E[Y|X \leq 1/2]\mathbf{P}(X \leq 1/2) + E[Y|X > 1/2]\mathbf{P}(X > 1/2) = \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{4}.$$

b An application of the total probability theorem, instead of the total expectation theorem, in Part a, shows that

$$f_Y(y) = f_{Y|X\leq 1/2}(y)\mathbf{P}(X \leq 1/2) + f_{Y|X>1/2}(y)\mathbf{P}(X > 1/2) = \text{uniform between 0 and 1/2.}$$

Therefore,

$$E\left[\frac{Y}{1-Y}\right] = \int_0^{1/2} 2\cdot\frac{y}{1-y}\,dy = \int_{1/2}^1 2\cdot\frac{1-v}{v}\,dv = 2\ \ln v\Big|_{1/2}^1 - 2\ v\Big|_{1/2}^1 = 2\ln 2 - 1 \approx 0.38,$$

where we used a change of variable $v = 1 - y$.

Problem 27. Erlang Random Variable. *(Ilya Pollak and Bin Ni.)*
A random variable X has probability density

$$f_X(x) = \begin{cases} Cx^2e^{-\lambda x}, & x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where λ is a positive real number. (This is an example of an Erlang distribution.) Find

(a) The constant C.

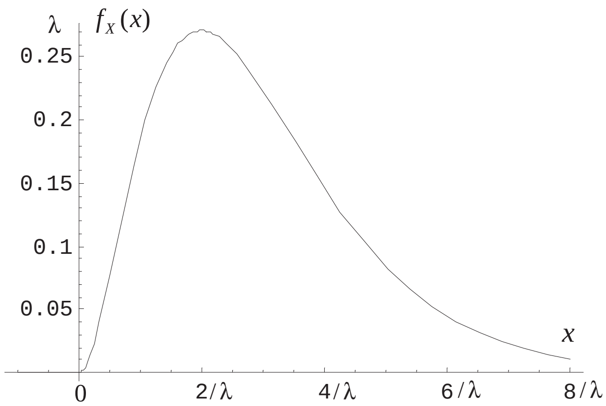
Solution. Because of the normalization axiom, the PDF should integrate to 1. On the other hand, we showed in class that the second moment of an exponential random variable with parameter λ is equal to $2/\lambda^2$–in other words, $\int_0^\infty x^2\lambda e^{-\lambda x}$ is equal to $2/\lambda^2$. We therefore have:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_X(x)dx = \frac{C}{\lambda}\int_0^\infty x^2\lambda e^{-\lambda x}dx \\ &= \frac{C}{\lambda}\cdot\frac{2}{\lambda^2} \\ \Rightarrow C &= \frac{\lambda^3}{2}. \end{aligned}$$

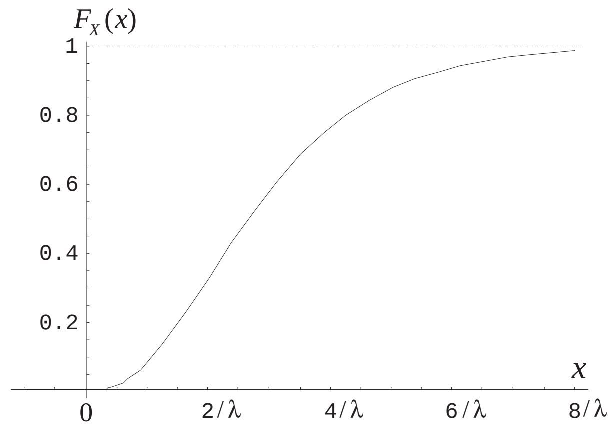
(b) The cumulative distribution function of X.

Solution. By definition,

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$



(a) PDF of X



(b) CDF of X

Figure 13: PDF and CDF of X in Problem 27.

From the expression of PDF we can see that if $x < 0$, the integral is equal to 0. When $x \geq 0$, we use integration by parts twice:

$$\begin{aligned}
F_X(x) &= \frac{\lambda^3}{2}\int_0^x u^2 e^{-\lambda u} du \\
&= \frac{\lambda^3}{2}\int_0^x u^2 d\left(-\frac{1}{\lambda}e^{-\lambda u}\right) \\
&= \frac{\lambda^3}{2}\left[-u^2\frac{1}{\lambda}e^{-\lambda u}\Big|_0^x + \int_0^x \frac{1}{\lambda}e^{-\lambda u} d(u^2)\right] \\
&= \frac{\lambda^3}{2}\left[-x^2\frac{1}{\lambda}e^{-\lambda x} + \frac{2}{\lambda}\int_0^x u d\left(-\frac{1}{\lambda}e^{-\lambda u}\right)\right] \\
&= \frac{\lambda^3}{2}\left[-x^2\frac{1}{\lambda}e^{-\lambda x} - \frac{2}{\lambda}\, u\frac{1}{\lambda}e^{-\lambda u}\Big|_0^x + \frac{2}{\lambda}\int_0^x \frac{1}{\lambda}e^{-\lambda u} du\right] \\
&= \frac{\lambda^3}{2}\left[-x^2\frac{1}{\lambda}e^{-\lambda x} - \frac{2}{\lambda^2}xe^{-\lambda x} - \frac{2}{\lambda^2}\,\frac{1}{\lambda}e^{-\lambda u}\Big|_0^x\right] \\
&= \frac{\lambda^3}{2}\left[\frac{2}{\lambda^3} - \frac{2 + 2\lambda x + \lambda^2 x^2}{\lambda^3}e^{-\lambda x}\right] \\
&= 1 - \left(1 + \lambda x + \frac{\lambda^2 x^2}{2}\right)e^{-\lambda x}.
\end{aligned}$$

The PDF and CDF are depicted in Figure 13.

(c) The probability $\mathbf{P}(0 \leq X \leq 1/\lambda)$.

Solution. By definition of CDF, we have:

$$\mathbf{P}(0 \leq X \leq 1/\lambda) = F_X(1/\lambda) - F_X(0).$$

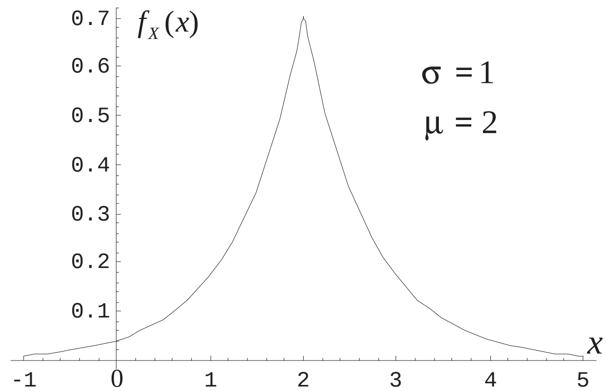


Figure 14: Laplace PDF.

By plugging in $1/\lambda$ and 0 into the expression of CDF obtained in Part(b), we get:

$$\begin{aligned} F_X(1/\lambda) &= 1-(1+1+1/2)e^{-1} = 1-\frac{5}{2e}, \\ F_X(0) &= 0, \\ \mathbf{P}(0 \le X \le 1/\lambda) &= F_X(1/\lambda) = 1-\frac{5}{2e} \approx 0.08030. \end{aligned}$$

Problem 28. LAPLACE RANDOM VARIABLE. *(Ilya Pollak and Bin Ni.)*
A random variable X has probability density

$$f_X(x) = Ce^{-\frac{\sqrt{2}|x-\mu|}{\sigma}},$$

where μ is a real number and σ is a positive real number. (This is a Laplace or double-sided exponential distribution.) Find

(a) The constant C.

Solution. Applying the normalization axiom,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_X(x)dx = C\int_{-\infty}^{\infty} e^{-\frac{\sqrt{2}|x-\mu|}{\sigma}}dx \\ &= 2C\int_{\mu}^{\infty} e^{-\frac{\sqrt{2}(x-\mu)}{\sigma}}dx \qquad \text{(due to symmetry about } \mu) \\ &= \sqrt{2}\sigma C\int_0^{\infty} e^{-u}du \qquad (u = \frac{\sqrt{2}(x-\mu)}{\sigma}) \\ &= \sqrt{2}\sigma C \\ \Rightarrow C &= \frac{1}{\sqrt{2}\sigma}. \end{aligned}$$

(b) The mean of X.

Solution. Applying the definition of expectation,

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} x C e^{-\frac{\sqrt{2}|x-\mu|}{\sigma}} dx.$$

Making a change of variable $u = x - \mu$, we have:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} (u+\mu) C e^{-\frac{\sqrt{2}|u|}{\sigma}} du \\ &= \underbrace{C \int_{-\infty}^{\infty} u e^{-\frac{\sqrt{2}|u|}{\sigma}} du}_{0} + \mu \underbrace{\int_{-\infty}^{\infty} C e^{-\frac{\sqrt{2}|u|}{\sigma}} du}_{1} \\ &= \mu. \end{aligned}$$

(c) The variance of X.

Solution. Applying the definition of variance,

$$\begin{aligned} \mathrm{var}[X] &= \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx \\ &= \frac{1}{\sqrt{2}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{\sqrt{2}|x-\mu|}{\sigma}} dx \\ &= \frac{1}{2} \int_{-\infty}^{\infty} u^2 \frac{\sqrt{2}}{\sigma} e^{-\frac{\sqrt{2}|u|}{\sigma}} du \\ &= \int_{0}^{\infty} u^2 \frac{\sqrt{2}}{\sigma} e^{-\frac{\sqrt{2}u}{\sigma}} du \\ &= \text{the second moment of an exponential r.v. with } \lambda = \frac{\sqrt{2}}{\sigma} \\ &= \frac{2}{\left(\frac{\sqrt{2}}{\sigma}\right)^2} = \sigma^2. \end{aligned}$$

Problem 29. Let X be a normal random variable with mean μ and variance σ^2.

(a) Compute the probabilities

$$\begin{aligned} \mathbf{P}(X \le \mu + \sigma) \\ \mathbf{P}(X \le \mu - \sigma) \\ \mathbf{P}(X \le \mu + 2\sigma) \end{aligned}$$

Solution. *(Ilya Pollak and Bin Ni.)*
If we define random variable $Y = (X - \mu)/\sigma$, then Y is a standard normal random variable—i.e.,

a normal random variable with zero mean and unit variance.

$$\begin{aligned}
\mathbf{P}(X \le \mu+\sigma) &= \mathbf{P}\left(\frac{X-\mu}{\sigma} \le 1\right) = \mathbf{P}(Y \le 1) = \Phi(1) \approx 0.8413, \\
\mathbf{P}(X \le \mu-\sigma) &= \mathbf{P}\left(\frac{X-\mu}{\sigma} \le -1\right) = \mathbf{P}(Y \le -1) = \Phi(-1) = 1-\Phi(1) \approx 0.1587, \\
\mathbf{P}(X \le \mu+2\sigma) &= \mathbf{P}\left(\frac{X-\mu}{\sigma} \le 2\right) = \mathbf{P}(Y \le 2) = \Phi(2) \approx 0.9772.
\end{aligned}$$

(b) Compute the probabilities

$$\mathbf{P}(\mu-\sigma \le X \le \mu+\sigma)$$
$$\mathbf{P}(\mu-2\sigma \le X \le \mu+2\sigma)$$

Solution

$$\begin{aligned}
\mathbf{P}(\mu-\sigma \le X \le \mu+\sigma) &= \mathbf{P}(X \le \mu+\sigma) - \mathbf{P}(X \le \mu-\sigma) \\
&\approx 0.8413 - 0.1587 = 0.6826, \\
\mathbf{P}(\mu-2\sigma \le X \le \mu+2\sigma) &= \mathbf{P}(X \le \mu+2\sigma) - \mathbf{P}(X \le \mu-2\sigma) \\
&\approx 0.9772 - (1-0.9772) = 0.9544.
\end{aligned}$$

Problem 30. *(Drake [2], problem 2.17. Solutions by Ilya Pollak and Bin Ni.)*
Stations A and B are connected by two *parallel* message channels. A message from A to B is sent over both channels at the same time. Continuous random variables X and Y represent the message delays (in hours) over parallel channels I and II, respectively. These two random variables are independent, and both are uniformly distributed from 0 to 1 hours.

A message is considered "received" as soon as it arrives on any one channel, and it is considered "verified" as soon as it has arrived over both channels.

(a) Determine the probability that a message is received within 15 minutes after it is sent.

Solution. Because the marginal PDF's are uniform and the X, Y are independent, the pair (X, Y) is uniformly distributed in the unit square as depicted in Figure 15. The probability of any event in this sample space is equal to its area. The event that "a message is received within 15 minute" is equivalent to "$X \le 1/4$ or $Y \le 1/4$", which corresponds to the shaded region in Figure 15(a). The probability of this event is equal to the area which is 7/16.

(b) Determine the probability that the message is received but not verified within 15 minutes after it is sent.

Solution. This event is equivalent to "($X \le 1/4$ and $Y > 1/4$) or ($Y \le 1/4$ and $X > 1/4$)", which is depicted as the shaded region in Figure 15 (b). The probability of this event is equal to 3/16+3/16=3/8.

(c) Let T represent the time (in hours) between transmission at A and verification at B. Determine the cumulative distribution function $F_T(t)$, and then differentiate it to obtain the PDF $f_T(t)$.

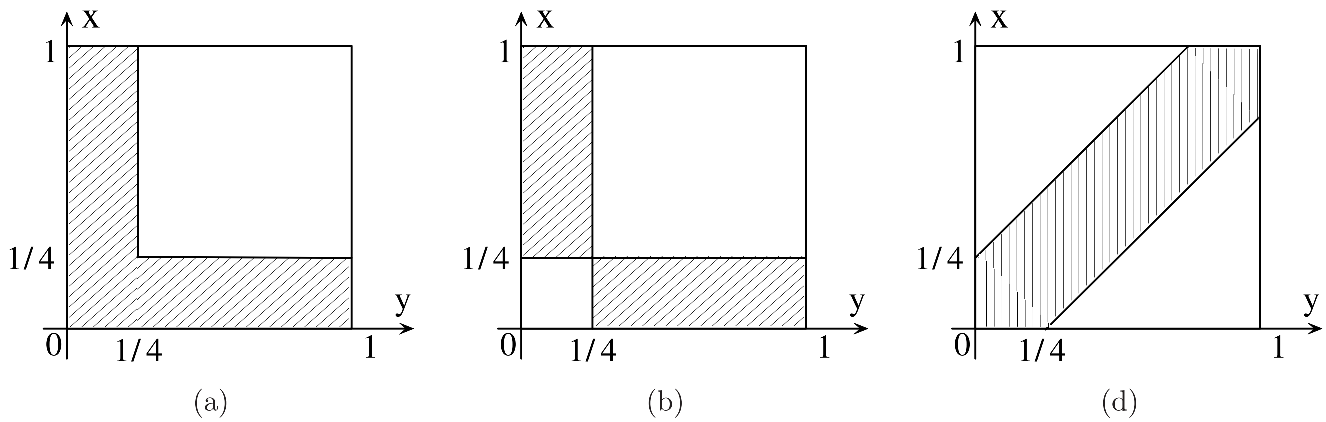


Figure 15: The events for Parts (a), (b), and (d) for Problem 30. The joint PDF of X and Y is uniform over the unit square.

Solution. By definition of CDF,

$$\begin{aligned} F_T(t) &= \mathbf{P}(T \le t) = \mathbf{P}(X \le t \text{ and } Y \le t) = \mathbf{P}(X \le t)\mathbf{P}(Y \le t) \qquad \text{(due to independence)} \\ &= F_X(t)F_Y(t) = \begin{cases} 0, & t < 0, \\ t^2, & 0 \le t \le 1, \\ 1, & t > 1. \end{cases} \end{aligned}$$

Differentiating, we get:

$$f_T(t) = \frac{dF_T(t)}{dt} = \begin{cases} 2t, & 0 \le t \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(d) If the attendant at B goes home 15 minutes after the message is received, what is the probability that he is present when the message should be verified?

Solution. If he is present for verification, it means that the message arrived over one channel within 15 minutes of arriving over the other one: $|X - Y| \le 1/4$. This is illustrated in Figure 15(d). The probability of this event is equal to 1 minus the area of the two triangles and we get:

$$\mathbf{P}(|X - Y| \le 1/4) = 1 - (1 - 1/4)^2 = 7/16.$$

(e) If the attendant at B leaves for a 15-minute coffee break right after the message is received, what is the probability that he is present at the proper time for verification?

Solution. This event is equivalent to $|X - Y| > 1/4$. Its probability is:

$$1 - \mathbf{P}(|X - Y| \le 1/4) = 9/16.$$

(f) The management wishes to have the maximum probability of having the attendant present at *both* reception and verification. Would they do better to let him take his coffee break as described above or simply allow him to go home 45 minutes after transmission?

Solution. If the attendant is allowed to go home 45 minutes after the transmission, the probability that he will be present at the both reception and verification is equal to the probability that both messages arrive within 45 minutes. That is:

$$\mathbf{P}(X \leq 3/4 \text{ and } Y \leq 3/4) = (3/4)^2 = 9/16.$$

This probability is equal to what we obtained in Part (e). So it does not matter which decision the manager will make in terms of maximizing the probability of the attendant's presence at both reception and verification. (But note that the coffee break strategy will keep the attendant at work for a longer time on average. Indeed, when the message is received more than 45 minutes after the transmission, he will need to stay there for more than 45 minutes. In the second scheme, his working time is fixed to be 45 minutes.)

Problem 31. *(Drake [2], Section 2-15, Example 2.)*
Each day he leaves home for the local casino, Oscar spins a biased wheel of fortune to determine how much money to take with him. He takes exactly X hundred dollars with him, where X is a continuous random variable described by the following probability density function:

$$f_X(x) = \begin{cases} \frac{x}{8}, & 0 \leq x \leq 4 \\ 0, & \text{otherwise,} \end{cases}$$

where x is in hundreds of dollars. As a matter of convenience, we are assuming that the currency is infinitely divisible. (Rounding off to the nearest penny wouldn't matter much.)

Oscar has a lot of experience at this. All of it is bad. Decades of experience have shown that, over the course of an evening, Oscar never wins. In fact, the amount with which he returns home on any particular night is uniformly distributed between zero and the amount with which he started out.

Let the random variable Y represent the amount (in hundreds of dollars) Oscar brings home on any particular night.

(a) Determine $f_{X,Y}(x, y)$, the joint PDF for his original wealth X and his terminal wealth Y on any night. Are X and Y independent? Present a convincing argument for your answer.

(b) Determine and plot $f_Y(y)$, the marginal PDF for the amount Oscar will bring home on any night.

(c) Determine the expected value of Oscar's loss on any particular night.

(d) On one particular night, we learn that Oscar returned home with less than \$200. For that night, determine the conditional probability of each of the following events:

(i) He started out for the casino with less than \$200.

(ii) His loss was less than \$100.

(iii) His loss was exactly \$75.

Solution.

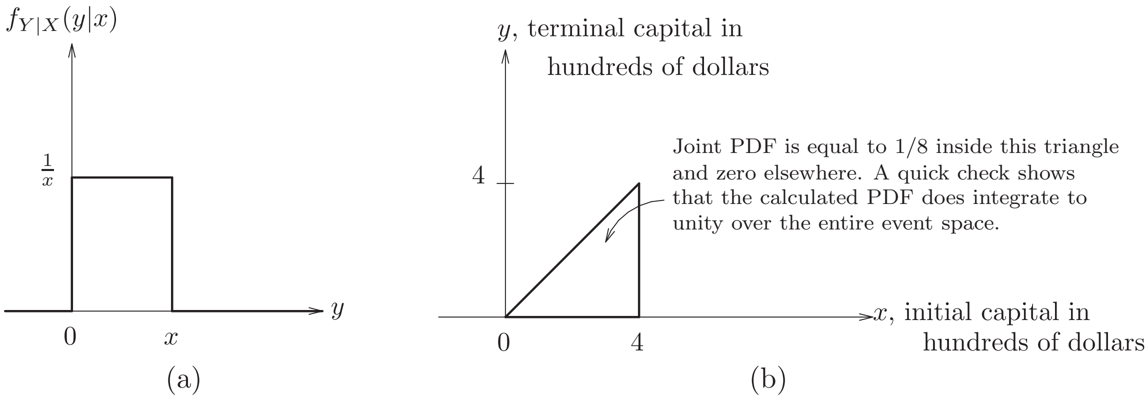


Figure 16: (a) Conditional PDF of Y given X. (b) The support of the joint PDF.

(a) From the problem statement we obtain:

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{x}, & 0 \le y \le x \le 4 \\ 0, & \text{otherwise} \end{cases}$$

This conditional density is shown in Fig. 16. This, in conjunction with the given $f_X(x)$, is used to determine the joint PDF:

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x) = \begin{cases} \frac{x}{8} \cdot \frac{1}{x} = \frac{1}{8}, & 0 \le y \le x \le 4, \\ 0, & \text{otherwise.} \end{cases}$$

This result may be displayed in an x, y event space as shown in Fig. 16(b). If X and Y were independent, then the conditional density $f_{Y|X}(y|x)$ would be equal to the marginal density $f_Y(y)$ of Y, and would not depend on the value x. It is, however, given that the conditional density of Y given $X = x$ is different for different values of x (specifically, it is uniform between zero and x). Therefore, X and Y are not independent.

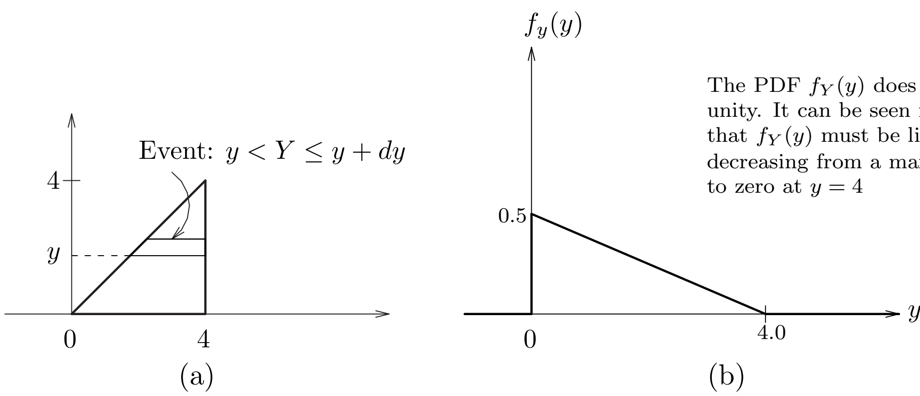


Figure 17: (a) To obtain the marginal PDF of Y, integrate the joint PDF along the x-axis over the triangular region. (b) The marginal PDF of Y.

(b) For $0 \le y \le 4$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{x=y}^{4} \frac{1}{8} dx = \frac{1}{8}(4 - y).$$

This calculation and its result are illustrated in Fig. 17. Notice that the product of the marginal densities of X and Y is not equal to the joint density of X and Y which was computed in Part (a). This is another way of seeing that X and Y are not independent.

(c)

$$E[X-Y] = \int_y \int_x (x-y) f_{X,Y}(x,y)\,dx\,dy$$

We must always be careful of the limits on the integrals when we substitute actual expressions for the joint PDF. We will integrate over x first.

$$\begin{aligned}
E[X-Y] &= \int_{y=0}^{4}\int_{x=y}^{4} \frac{1}{8}(x-y)\,dx\,dy = \int_{y=0}^{4}\left[\frac{1}{8}\left(\frac{x^2}{2}-xy\right)\right]_{x=y}^{x=4} dy \\
&= \int_0^4 \frac{1}{8}\left(8-4y-\frac{y^2}{2}+y^2\right)dy = \int_0^4\left(1-\frac{y}{2}+\frac{y^2}{16}\right)dy \\
&= \left[y-\frac{y^2}{4}+\frac{y^3}{48}\right]_0^4 = \frac{4}{3} = \$133.33.
\end{aligned}$$

Alternately, we can obtain $E[X-Y]$ as follows:

$$\begin{aligned}
E[X-Y] = E[X]-E[Y] &= \int_{x=-\infty}^{\infty} x f_X(x)\,dx - \int_{y=-\infty}^{\infty} y f_Y(y)\,dy \\
&= \int_0^4 x\cdot\frac{x}{8}\,dx - \int_0^4 y\cdot\frac{4-y}{8}\,dy \\
&= \$133.33
\end{aligned}$$

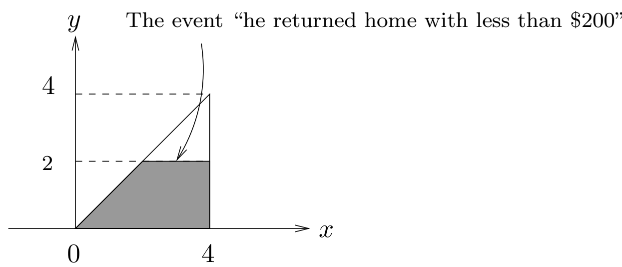


Figure 18: The conditional sample space for Part (d).

(d) Given that Oscar returned home with less than \$200, we work in the appropriate conditional sample space. At all points consistent with the conditioning event, the conditional joint PDF is equal to the original joint PDF scaled up by the reciprocal of the a priori probability of the conditioning event. Since the original joint PDF is uniform, this means that the conditional joint PDF is also uniform. The area of the conditioning event is 6, therefore the conditional joint PDF given that Oscar returned with less than \$200 is equal to $\frac{1}{6}$ in the region where it is nonzero, i.e. the gray set in Fig. 18. Now we may answer all the questions in this conditional space.

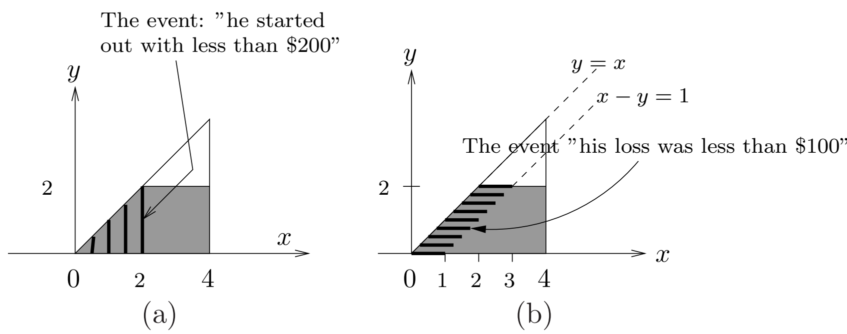


Figure 19: In both figures, the conditioning event "returned home with less than $200" is the gray set. The striped events are: (a) "started out with less than $200" and (b) "lost less than $100."

(i) See Fig. 19(a). The gray event is the conditioning event "returned home with less than $200", and the striped event is the event "started out with less than $200." Since the conditional joint PDF is uniform, we can find the conditional probability as the ratio of the striped area to the gray area which is $2/6 = 1/3$.

(ii) See Fig. 19(b). Similarly, the conditional probability is the ratio of the striped area to the gray area, $2/6 = 1/3$. We realize that, in general, we would have to integrate the conditional PDF over the appropriate events to obtain their probabilities. Only because the conditional PDF is constant have we been able to reduce the integrations to simple ratios of areas.

(iii) As long as we allow the currency to be infinitely divisible, the conditional probability measure associated with the event $X - Y = 75$ is equal to zero. The integral of the joint PDF $f_{X,Y}(x, y)$ over the line representing this event in the x, y event space is equal to zero.

Problem 32. *(Solutions by Ilya Pollak and Bin Ni.)*
Continuous random variables X and Y have the following joint probability density function:

$$f_{X,Y}(x, y) = \begin{cases} c, & 0 \le y \le 2x \le 4 \\ 0, & \text{otherwise,} \end{cases}$$

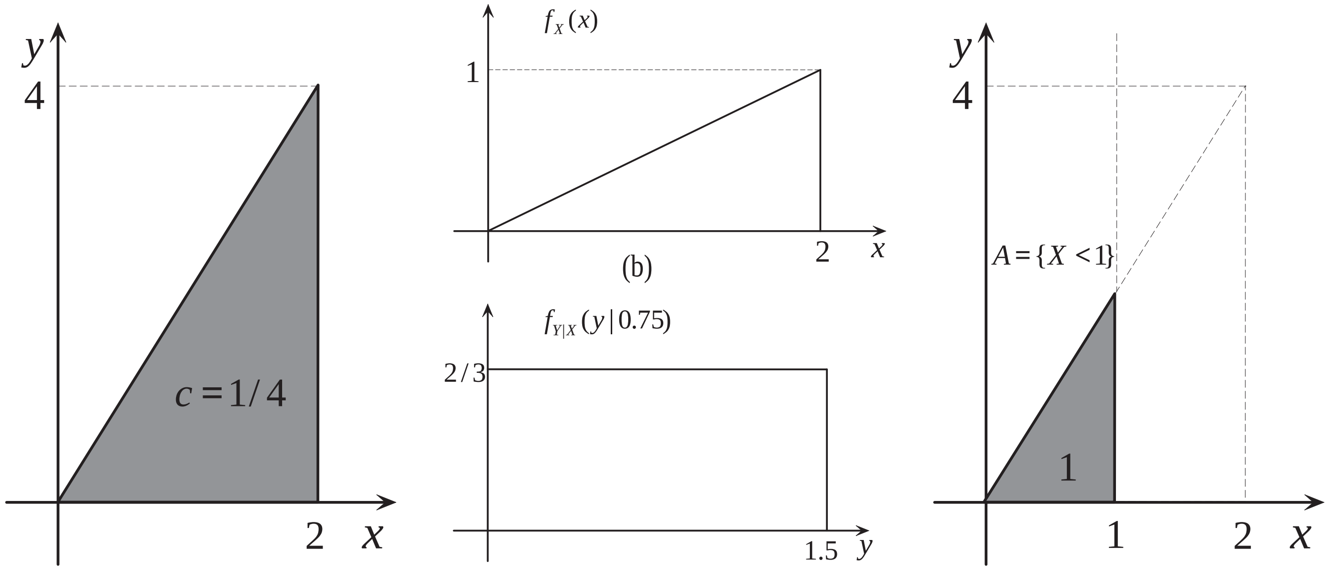
where c is a constant.

(a) Are X and Y independent? Present a convincing argument for your answer.

Solution. The joint PDF $f_{X,Y}(x, y)$ is depicted in Fig. 20(a). We can see that X and Y are not independent, because the conditional PDF $p_{X|Y}(x|y_0)$, which is the normalized version of the slice of $f_{X,Y}(x, y)$ at $y = y_0$, does depend on the value of y_0. For example, when $y_0 = 0$, X is uniformly distributed on [0,2] whereas when $y_0 = 2$, the range of X becomes [1,2].

(b) Prepare fully labeled plots of $f_X(x)$ and $f_{Y|X}(y|0.75)$.

Solution. The joint PDF is uniform over the dark triangle in Fig. 20(a). In order for it to be correctly normalized (i.e., integrate to one), the constant c must be equal to 1/(area of the triangle) $= 1/4$.



(a) Support of $f_{X,Y}(x, y)$.
(c)
(d) Support of $f_{X,Y|A}(x, y)$.

Figure 20: Problem 32.

It is obvious that $f_X(x) = 0$ for $x < 0$, $x > 2$. To get $f_X(x_0)$ at a point x_0 which is between 0 and 2, we need to integrate $f_{X,Y}(x_0, y)$ with respect to y. In other words, we need to integrate the constant $c = 1/4$ over the vertical straight line segment from $(x_0, 0)$ to $(x_0, 2x_0)$: $f_X(x_0) = (1/4)2x_0 = x_0/2$.

An alternative method is to notice that the length of the line segment over which we are integrating, grows linearly with x for $0 \le x \le 2$, and therefore $f_X(x)$ is a linear function for $0 \le x \le 2$. Its slope is found by noticing that $1 = \int f_X(x)dx$ which must be equal to the area of the triangle in Fig. 20(b), which is $0.5 f_X(2) \cdot 2$, and therefore $f_X(2) = 1$.

In summary, as shown in Fig. 20(b), we have:

$$f_X(x) = \begin{cases} \frac{1}{2}x & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

From Fig. 20(a), we note that, given $X = 0.75$, Y is uniform between 0 and 1.5:

$$f_{Y|X}(y|0.75) = \begin{cases} \frac{2}{3} & 0 \le y \le 1.5 \\ 0 & \text{otherwise,} \end{cases}$$

which is depicted in Fig. 20(c).

(c) Evaluate $E[X|Y = 3]$.

Solution. Given $Y = 3$, we can see that X is uniformly distributed on [1.5, 2]. Therefore the expectation is just the center of the distribution, which equals 1.75.

(d) Let $R = XY$ and let A be the event $X < 1$. Evaluate $E[R|A]$.

Solution. By definition of expectation, we get:

$$E[R|A] = E[XY|A] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y|A}(x,y)dydx. \tag{20}$$

The conditional joint PDF, $f_{X,Y|A}(x,y)$ can be obtained by restricting the joint PDF $f_{X,Y}(x,y)$ to A and re-normalizing–i.e., it is uniform over A, as shown in Fig. 20(d):

$$f_{X,Y|A}(x,y) = \begin{cases} 1, & 0 \le y \le 2x < 2 \\ 0, & \text{otherwise.} \end{cases}$$

Substitute this into Eq. 20 to get:

$$E[R|A] = \int_0^1 \int_0^{2x} xydydx = \int_0^1 x \left.\frac{y^2}{2}\right|_0^{2x} dx = \int_0^1 2x^3 dx = 2\left.\frac{x^4}{4}\right|_0^1 = \frac{1}{2}.$$

Problem 33. *(Drake [2], problem 2.26. Solutions by Ilya Pollak and Bin Ni.)*
Melvin Fooch, a student of probability theory, has found that the hours he spends working (W) and sleeping (S) in preparation for a final exam are random variables described by:

$$f_{W,S}(w,s) = \begin{cases} K & \text{if } 10 \le w+s \le 20,\ w \ge 0,\ s \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

What poor Melvin does not know, and even his best friends will not tell him, is that working only furthers his confusion and that his grade, G, can be described by

$$G = 2.5(S - W) + 50.$$

(a) Evaluate constant K.

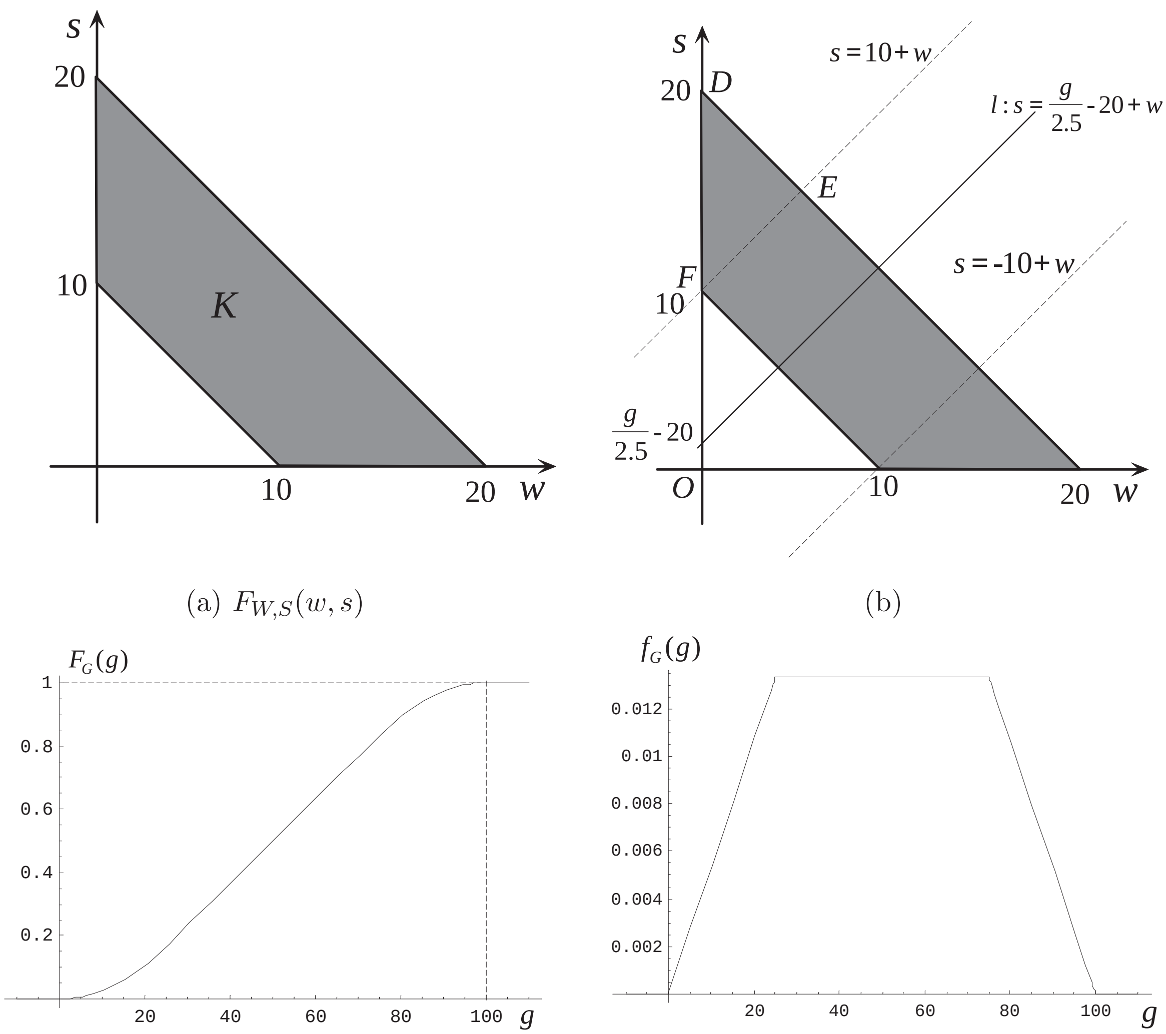
Solution. The joint PDF is depicted in Fig. 21(a). The value of the function is constant K in the dark region, and 0 elsewhere. Hence the integration of $f_{W,S}(w,s)$ over the whole plane equals $K \cdot$ Area of dark region. Since the joint PDF must integrate to 1, we have:

$$\begin{aligned} K \cdot \text{Area of dark region} &= 1 \\ \Rightarrow K &= 1/\text{Area of dark region} \\ &= 1/(\text{Area of big triangle} - \text{Area of small triangle}) \\ &= 1/(200-50) = 1/150. \end{aligned}$$

(b) The instructor has decided to pass Melvin if, on the exam, he achieves $G \ge 75$. What is the probability that this will occur?

Solution. By definition of G, we have:

$$\begin{aligned} \mathbf{P}(G > 75) &= \mathbf{P}(2.5(S-W)+50 > 75) \\ &= \mathbf{P}(2.5(S-W) > 25) = \mathbf{P}(S - W > 10) \\ &= \mathbf{P}(S > 10 + W). \end{aligned}$$



(a) $F_{W,S}(w,s)$

(b)

(c) $F_G(g)$

(d) $f_G(g)$

Figure 21: Illustrations for Problem 3.

The event $A = \{S > 10 + W\}$ corresponds to the semi-infinite plane above the upper dashed line in Fig. 21(b). By definition of PDF, we have:

$$\begin{aligned}
\mathbf{P}(A) &= \iint_{(s,w)\in A} f_{W,S}(w,s)dwds \\
&= K \cdot \text{Area of DEF} = \frac{1}{150} \cdot 25 = \frac{1}{6}
\end{aligned}$$

(c) Make a neat and fully labeled sketch of the probability density function $f_G(g)$.

Solution. Given the range of W and S, it's not hard to check that G ranges from 0 to 100. Therefore $f_G(g) = 0$ for $g < 0, g > 100$. For $0 \leq g \leq 100$, it will be easier to get the CDF first and then take derivative. By definition of CDF, we have:

$$\begin{aligned}
F_G(g) &= \mathbf{P}(G < g) = \mathbf{P}\left(S - W < \frac{g}{2.5} - 20\right) && (21) \\
&= \mathbf{P}\left(S < \frac{g}{2.5} - 20 + W\right). && (22)
\end{aligned}$$

Similarly to the last part, $\mathbf{P}\left(S < \frac{g}{2.5} - 20 + W\right)$ equals K times the area of the piece of the dark region below the straight line $l : s < \frac{g}{2.5} - 20 + w$ which is depicted in Fig. 21(b). Because of the shape of the dark region, Eq. 22 needs to be evaluated under three cases:

Case 1. $-20 \leq \frac{g}{2.5} - 20 < -10$, which is equivalent to $0 \leq g < 25$:
In this case, l is below the lower dashed line in the figure, and so the area of interest is:

$$A(g) = \frac{1}{4}\left(\frac{g}{2.5}\right) = \frac{g^2}{25},$$

and therefore

$$F_G(g) = KA(g) = \frac{g^2}{3750}.$$

Case 2. $-10 \leq \frac{g}{2.5} - 20 < 10$, which is equivalent to $25 \leq g < 75$:
In this case, l is between the lower dashed line and the upper dashed line in the figure, and so

$$A(g) = 5\left(\frac{g}{2.5} - 10\right) + 25 = 2g - 25,$$

and

$$F_G(g) = KA(g) = \frac{2g - 25}{150}.$$

Case 3. $10 \leq \frac{g}{2.5} - 20 \leq 20$, which is equivalent to $75 \leq g \leq 100$:
In this case, l is above the upper dashed line in the figure, and so

$$A(g) = 150 - \frac{1}{4}\left(40 - \frac{g}{2.5}\right)^2 = 8g - \frac{g^2}{25} - 250,$$

and

$$F_G(g) = KA(g) = \frac{8g - g^2/25 - 250}{150}.$$

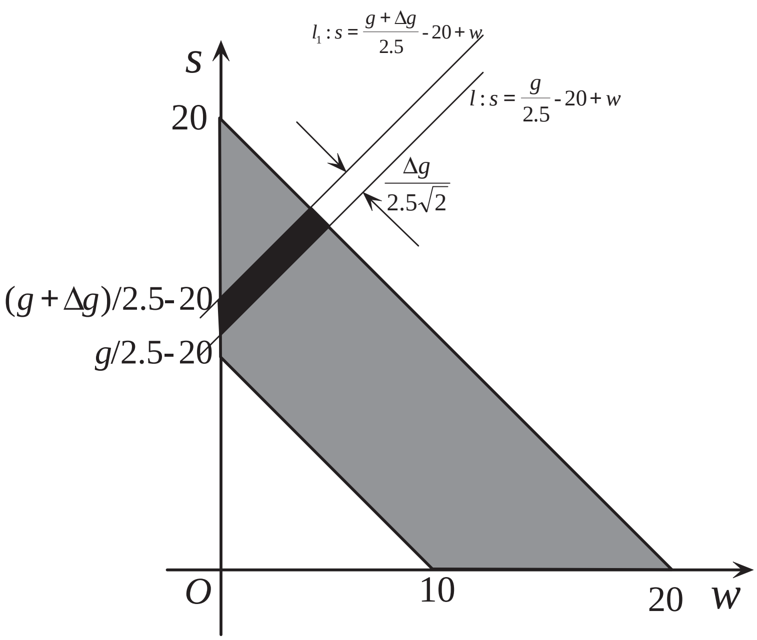


Figure 22: Problem 3(c), alternative method.

Combining the three cases, $F_G(g)$ is depicted as in Fig. 21(c). If we differentiate, we'll get:

$$f_G(g) = \begin{cases} g/1875 & 0 \le g < 25 \\ 1/75 & 25 \le g < 75 \\ 4/75 - g/1875 & 75 \le g \le 100 \\ 0 & \text{otherwise,} \end{cases}$$

which is depicted in Figure 21(d).

Another way to solve this problem is based on the following fact:

$$f_G(g) = \frac{dF_G(g)}{dg} = \lim_{\Delta g \to 0} \frac{F_G(g+\Delta g) - F_G(g)}{\Delta g}.$$

As depicted in Fig. 22, $F_G(g+\Delta g) - F_G(g)$ is equal to the probability of the black region, which is the part of the dark region between straight line l and $l_1 : s = \frac{g+\Delta g}{2.5} - 20 + w$. When Δg is small, the area of the black region can be approximated by $\frac{\Delta g}{2.5\sqrt{2}}$ times the length of the segment of line l that is cut by the dark region. Therefore we have:

$$f_G(g) = K \frac{1}{2.5\sqrt{2}} \cdot \text{length of the segment of line } l \text{ that is cut by the dark region}$$

When $0 \le g \le 25$ it is linearly increasing, when $25 \le g \le 75$ it is constant, and when $75 \le g \le 100$ it is linearly decreasing. It is therefore a trapezoid, as shown in Fig. 21(d). The lengths of the two parallel sides are 100, 50 respectively. The height of the trapezoid equals:

$$f_G(25) = K\frac{1}{2.5\sqrt{2}} \cdot \frac{10}{\sqrt{2}} = \frac{1}{75}$$

It can also be found according to the normalization axiom:

$$\begin{aligned} \frac{100+50}{2} f_G(25) &= 1 \\ \Rightarrow f_G(25) &= \frac{1}{75}. \end{aligned}$$

(d) Melvin, true to form, got a grade of exactly 75 on the exam. Determine the conditional probability that he spent less than one hour working in preparation for this exam.

Solution. Given that he got a 75, the set of possible outcomes is just segment EF in Fig. 21(b). Since the joint PDF is constant on the set, the conditional density of W is uniform on [0,5], and $\mathbf{P}(W < 1 | G = 75) = 1/5$.

References

[1] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to Probability*, 2nd Edition, Athena Scientific, Belmont, Massachusetts, 2008.

[2] A. Drake. *Fundamentals of Applied Probability Theory*, McGraw-Hill, Inc., 1988.

[3] A. Leon-Garcia. *Probability and Random Processes for Electrical Engineering*, 2nd Edition, Addison Wesley Longman, 1994.

[4] J. Montier. *Behavioural Investing: A Practitioners Guide to Applying Behavioural Finance.* Wiley, 2007.

[5] Q.F. Stout and B. Warren. Tree algorithms for unbiased coin tossing with a biased coin. *The Annals of Probability*, 12(1):212–222, Feb. 1984.