Example 5.3. Upper Bounds in the Chebyshev Inequality. When X is known to take values in a range $[a, b]$, we claim that $\sigma^2 \leq (b-a)^2/4$. Thus, if σ^2 is unknown, we may use the bound $(b-a)^2/4$ in place of σ^2 in the Chebyshev inequality, and obtain

$$\mathbf{P}\big(|X-\mu| \geq c\big) \leq \frac{(b-a)^2}{4c^2}, \qquad \text{for all } c > 0.$$

To verify our claim, note that for any constant γ, we have

$$\mathbf{E}\big[(X-\gamma)^2\big] = \mathbf{E}[X^2] - 2\mathbf{E}[X]\gamma + \gamma^2,$$

and the above quadratic is minimized when $\gamma = \mathbf{E}[X]$. It follows that

$$\sigma^2 = \mathbf{E}\Big[\big(X - \mathbf{E}[X]\big)^2\Big] \leq \mathbf{E}\big[(X-\gamma)^2\big], \qquad \text{for all } \gamma.$$

By letting $\gamma = (a+b)/2$, we obtain

$$\sigma^2 \leq \mathbf{E}\left[\left(X - \frac{a+b}{2}\right)^2\right] = \mathbf{E}\big[(X-a)(X-b)\big] + \frac{(b-a)^2}{4} \leq \frac{(b-a)^2}{4},$$

where the equality above is verified by straightforward calculation, and the last inequality follows from the fact

$$(x-a)(x-b) \leq 0$$

for all x in the range $[a, b]$.

The bound $\sigma^2 \leq (b-a)^2/4$ may be quite conservative, but in the absence of further information about X, it cannot be improved. It is satisfied with equality when X is the random variable that takes the two extreme values a and b with equal probability $1/2$.

Problem 3.* **Jensen inequality.** A twice differentiable real-valued function f of a single variable is called **convex** if its second derivative $(d^2 f/dx^2)(x)$ is nonnegative for all x in its domain of definition.

(a) Show that the functions $f(x) = e^{\alpha x}$, $f(x) = -\ln x$, and $f(x) = x^4$ are all convex.

(b) Show that if f is twice differentiable and convex, then the first order Taylor approximation of f is an underestimate of the function. that is,

$$f(a) + (x-a)\frac{df}{dx}(a) \le f(x).$$

for every a and x.

(c) Show that if f has the property in part (b), and if X is a random variable, then

$$f\big(\mathbf{E}[X]\big) \le \mathbf{E}\big[f(X)\big].$$

Solution. (a) We have

$$\frac{d^2}{dx^2}e^{ax} = a^2 e^{ax} > 0, \qquad \frac{d^2}{dx^2}(-\ln x) = \frac{1}{x^2} > 0, \qquad \frac{d^2}{dx^2}x^4 = 4\cdot 3\cdot x^2 \ge 0.$$

(b) Since the second derivative of f is nonnegative, its first derivative must be nondecreasing. Using the fundamental theorem of calculus, we obtain

$$f(x) = f(a) + \int_a^x \frac{df}{dt}(t)\,dt \ge f(a) + \int_a^x \frac{df}{dt}(a)\,dt = f(a) + (x-a)\frac{df}{dx}(a).$$

(c) Since the inequality from part (b) is assumed valid for every possible value x of the random variable X, we obtain

$$f(a) + (X-a)\frac{df}{dx}(a) \le f(X).$$

We now choose $a = \mathbf{E}[X]$ and take expectations, to obtain

$$f\big(\mathbf{E}[X]\big) + \big(\mathbf{E}[X] - \mathbf{E}[X]\big)\frac{df}{dx}\big(\mathbf{E}[X]\big) \le \mathbf{E}\big[f(X)\big],$$

or

$$f\big(\mathbf{E}[X]\big) \le \mathbf{E}\big[f(X)\big].$$

Solution to Problem 5.10. (a) Let $S_n = X_1 + \cdots + X_n$ be the total number of gadgets produced in n days. Note that the mean, variance, and standard deviation of S_n is $5n$, $9n$, and $3\sqrt{n}$, respectively. Thus,

$$\begin{aligned}
\mathbf{P}(S_{100} < 440) &= \mathbf{P}(S_{100} \le 439.5) \\
&= \mathbf{P}\left(\frac{S_{100} - 500}{30} < \frac{439.5 - 500}{30}\right) \\
&\approx \Phi(\frac{439.5 - 500}{30}) \\
&= \Phi(-2.02) \\
&= 1 - \Phi(2.02) \\
&= 1 - 0.9783 \\
&= 0.0217.
\end{aligned}$$

(b) The requirement $\mathbf{P}(S_n \ge 200 + 5n) \le 0.05$ translates to

$$\mathbf{P}\left(\frac{S_n - 5n}{3\sqrt{n}} \ge \frac{200}{3\sqrt{n}}\right) \le 0.05,$$

or, using a normal approximation,

$$1 - \Phi\left(\frac{200}{3\sqrt{n}}\right) \le 0.05,$$

and

$$\Phi\left(\frac{200}{3\sqrt{n}}\right) \ge 0.95.$$

From the normal tables, we obtain $\Phi(1.65) \approx 0.95$, and therefore,

$$\frac{200}{3\sqrt{n}} \ge 1.65,$$

which finally yields $n \le 1632$.

(c) The event $N \ge 220$ (it takes at least 220 days to exceed 1000 gadgets) is the same as the event $S_{219} \le 1000$ (no more than 1000 gadgets produced in the first 219 days). Thus,

$$\begin{aligned}
\mathbf{P}(N \ge 220) &= \mathbf{P}(S_{219} \le 1000) \\
&= \mathbf{P}\left(\frac{S_{219} - 5 \cdot 219}{3\sqrt{219}} \le \frac{1000 - 5 \cdot 219}{3\sqrt{219}}\right) \\
&= 1 - \Phi(2.14) \\
&= 1 - 0.9838 \\
&= 0.0162.
\end{aligned}$$

Solution to Problem 5.11. Note that W is the sample mean of 16 independent identically distributed random variables of the form $X_i - Y_i$, and a normal approximation is appropriate. The random variables $X_i - Y_i$ have zero mean, and variance equal to $2/12$. Therefore, the mean of W is zero, and its variance is $(2/12)/16 = 1/96$. Thus,

$$\mathbf{P}\big(|W| < 0.001\big) = \mathbf{P}\left(\frac{|W|}{\sqrt{1/96}} < \frac{0.001}{\sqrt{1/96}}\right) \approx \Phi\big(0.001\sqrt{96}\big) - \Phi\big(-0.001\sqrt{96}\big)$$

$$= 2\Phi\big(0.001\sqrt{96}\big) - 1 = 2\Phi(0.0098) - 1 \approx 2 \cdot 0.504 - 1 = 0.008.$$

Let us also point out a somewhat different approach that bypasses the need for the normal table. Let Z be a normal random variable with zero mean and standard deviation equal to $1/\sqrt{96}$. The standard deviation of Z, which is about 0.1, is much larger than 0.001. Thus, within the interval $[-0.001, 0.001]$, the PDF of Z is approximately constant. Using the formula $\mathbf{P}(z - \delta \le Z \le z + \delta) \approx f_Z(z) \cdot 2\delta$, with $z = 0$ and $\delta = 0.001$, we obtain

$$\mathbf{P}\big(|W| < 0.001\big) \approx \mathbf{P}(-0.001 \le Z \le 0.001) \approx f_Z(0) \cdot 0.002 = \frac{0.002}{\sqrt{2\pi}(1/\sqrt{96})} = 0.0078.$$

Example 9.3. Estimating the Parameter of an Exponential Random Variable. Customers arrive to a facility, with the ith customer arriving at time Y_i. We assume that the ith interarrival time, $X_i = Y_i - Y_{i-1}$ (with the convention $Y_0 = 0$) is exponentially distributed with unknown parameter θ, and that the random variables $X_1, \ldots, X_n$ are independent. (This is the Poisson arrivals model, studied in Chapter 6.) We wish to estimate the value of θ (interpreted as the arrival rate), on the basis of the observations $X_1, \ldots, X_n$.

The corresponding likelihood function is

$$f_X(x;\theta) = \prod_{i=1}^{n} f_{X_i}(x_i;\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i},$$

and the log-likelihood function is

$$\log f_X(x;\theta) = n \log\theta - \theta y_n,$$

where

$$y_n = \sum_{i=1}^{n} x_i.$$

The derivative with respect to θ is $(n/\theta) - y_n$, and by setting it to 0, we see that the maximum of $\log f_X(x;\theta)$, over $\theta \geq 0$, is attained at $\hat{\theta}_n = n/y_n$. The resulting estimator is

$$\hat{\Theta}_n = \left(\frac{Y_n}{n}\right)^{-1}.$$

It is the inverse of the sample mean of the interarrival times, and it can be interpreted as an empirical arrival rate.

Note that by the weak law of large numbers, Y_n/n converges in probability to $\mathbf{E}[X_i] = 1/\theta$, as $n \to \infty$. This can be used to show that $\hat{\Theta}_n$ converges to θ in probability, so the estimator is consistent.

Solution to Problem 9.4. (a) Figure 9.1 plots a mixture of two normal distributions. Denoting $\theta = (p_1, \mu_1, \sigma_1, \ldots, p_m, \mu_m, \sigma_m)$, the PDF of each X_i is

$$f_{X_i}(x_i;\theta) = \sum_{j=1}^{m} p_j \cdot \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{\frac{-(x_i - \mu_j)^2}{2\sigma_j^2}\right\}.$$

Using the independence assumption, the likelihood function is

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n} f_{X_i}(x_i;\theta) = \prod_{i=1}^{n}\left(\sum_{j=1}^{m} p_j \cdot \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{\frac{-(x_i - \mu_j)^2}{2\sigma_j^2}\right\}\right),$$

and the log-likelihood function is

$$\log f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{m} p_j \cdot \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{\frac{-(x_i-\mu_j)^2}{2\sigma_j^2}\right\}\right).$$

(b) The likelihood function is

$$p_1 \cdot \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right\} + (1-p_1) \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right\},$$

and is linear in p_1. The ML estimate of p_1 is

$$\hat{p}_1 = \begin{cases} 1, & \text{if } \dfrac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{\dfrac{-(x-\mu_1)^2}{2\sigma_1^2}\right\} > \dfrac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{\dfrac{-(x-\mu_2)^2}{2\sigma_2^2}\right\}, \\ 0, & \text{otherwise,} \end{cases}$$

and the ML estimate of p_2 is $\hat{p}_2 = 1 - \hat{p}_1$.

(c) The likelihood function is the sum of two terms [cf. the solution to part (b)], the involving μ_1, the second involving μ_2. Thus, we can maximize each term separ and find that the ML estimates are $\hat{\mu}_1 = \hat{\mu}_2 = x$.

(d) Fix $p_1, \ldots, p_m$ to some positive values. Fix $\mu_2, \ldots, \mu_m$ and $\sigma_2^2, \ldots, \sigma_m^2$ to s arbitrary (respectively, positive) values. If $\mu_1 = x_1$ and σ_1^2 tends to zero, the likeli $f_{X_1}(x_1;\theta)$ tends to infinity, and the likelihoods $f_{X_i}(x_i;\theta)$ of the remaining points (i are bounded below by a positive number. Therefore, the overall likelihood tenc infinity.

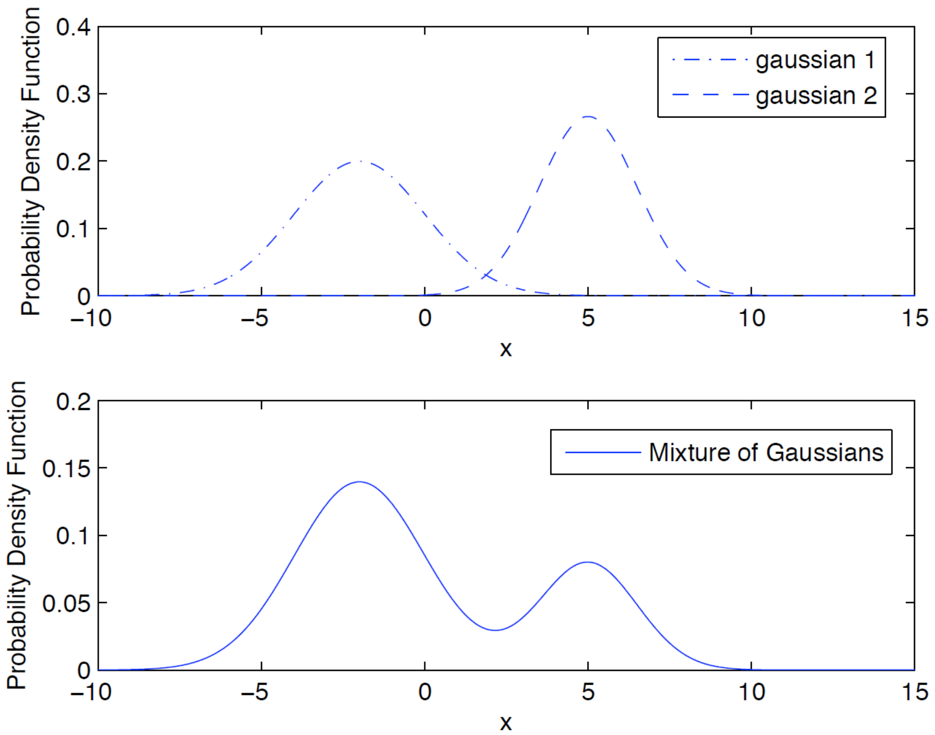

Figure 9.1: The mixture of two normal distributions with $p_1 = 0.7$ and $p_2 = 0.3$ in Problem 9.4.