Lecture 12: Quality Control I: Control of Location

10 October 2005

This lecture and the next will be about quality control methods. There are two reasons for this. First, it's intrinsically important for engineering, but the basic math is all stuff we've seen already — mostly, it's the central limit theorem. Second, it will give us a fore-taste of the issues which will come up in the second half of the course, on statistical inference. Your textbook puts quality control in the last chapter, but we'll go over it here instead.

1 Random vs. Systematic Errors

An important basic concept is to distinguish between two kinds of influences which can cause errors in a manufacturing process, or introduce variability into anything. On the one hand, there are **systematic** influences, which stem from particular factors, influence the process in a consistent way, and could be corrected by manipulating those factors. This is the mis-calibrated measuring instrument, defective materials, broken gears in a crucial machine, or the worker who decides to shut off the cement-stirer while he finishes his bottle of vodka. In quality control, these are sometimes called **assignable causes**. In principle, there is someone to blame for things going wrong, and you can find out who that is and stop them. On the other hand, some errors and variability are due to essentially **random**, **statistical** or **noise** factors, the ones with no consistent effect and where there's really no one to blame — it's just down to the fact that the world is immensely complicated, and no process will ever work itself out exactly the same way twice. No one's to blame for noise; once you've reduced it as far as is reasonable, you just have to learn to live with it.

The point of quality control methods is to help us tell the difference between systematic problems, where we want to figure out what's to blame, and noise, where asking "why did this have to happen to me?" is like a tossed coin asking "why did I have to come up heads?"

People are in general *very bad* at making this distinction on purely intuitve grounds, becaused we're much too inclined to see patterns, and agents responsible for those patterns, in pure noise. This has been extensively studied by psychologists, for instance. One classic set of experiments works like this. You

are set at a table on which there is a green light and red light, and you have to guess which one will come on next. You'll get a nickle for each correct guess. You are *not* told what pattern controls the lights. In fact there is no pattern, it's completely random, but biased, so that (say) the green light comes on 75% of the time. In this case, the way to maximize your reward is to always guess the more probable light.[1] People almost *never* do this, however; instead they come up with very elaborate rules for the sequence of lights, which get the proportions right, but do no better than chance at actually predicting the lights.[2] Interestingly, *rats* do just fine in this experiment; so do people who certain kinds of brain-damage, which prevent them from figuring out the pattern when there is one.[3]

The goal of quality control methods is to let ordinary engineers do as well as rats or people with brain-damage.

2 Control Charts for Location: Gaussian Process with Known Parameters

We are manufacturing something, say widgets, and want to make sure that it's got the right quantitative characteristic, say elasticity. To make everything easy on ourselves, we will start by assuming that, if everything is going according to plan — if the process is **in control** — this characteristic is a random variable, X, whose distribution is Gaussian with a *known* mean and standard deviation, $X \sim \mathcal{N}(\mu, \sigma)$. (We will see how to relax all these assumptions as we go along.) Every day, or hour, or month, as appropriate, we pick n a sample of n independent units of the output — n independent widgets — and look at the sample mean, $\overline{X}$. Because X is Gaussian, $\overline{X} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$ — it's got the same mean, but the standard deviation is smaller. Next, we plot each day's value for the sample mean vs. the day. Figure 1 shows what a typical result would look like, with $\mu = 12$, $\sigma = 3$, and $n = 4$; I chose 40 days. We expect that each day's output should be reasonably close to the mean, and that getting to far from the mean is a cause for alarm, but how far should things go before we get worried? There is a traditional answer, which goes back at least to the 1940s. It says to draw the

[1] Think of whether or not the green light comes on as a sequence of independent Bernoulli variables with success probability p. Then if you guess green with probability g, the probability of guessing correctly is $(1-g)(1-p) + gp = 1 - g - p + 2gp$. (The light is independent of your guess.) We want to maximize this with respect to g, so take the derivative:

$$\frac{\partial}{\partial g}(1 - g - p + 2gp) = 2p - 1$$

Since the derivative doesn't depend on g, the maximum is at the boundary — the upper one, $g = 1$, if $2p - 1 > 0$, i.e., if $p > 1/2$, and the lower one, $g = 0$, if $p < 1/2$. So you should either always guess green, if $p > 1/2$, or always guess red, if $p < 1/2$. If $p = 1/2$, it doesn't matter what you do.

[2] In terms of the previous footnote, they get $g = p$. This means that the fraction of the time they're correct is $(1-p)^2 + p^2$, which is sub-optimal (unless $p = 1/2$).

[3] George Wolford, Michael B. Miller and Michael Gazzaniga, "The Left Hemisphere's Role in Hypothesis Formation", *The Journal of Neuroscience* **20** (2000): RC64.

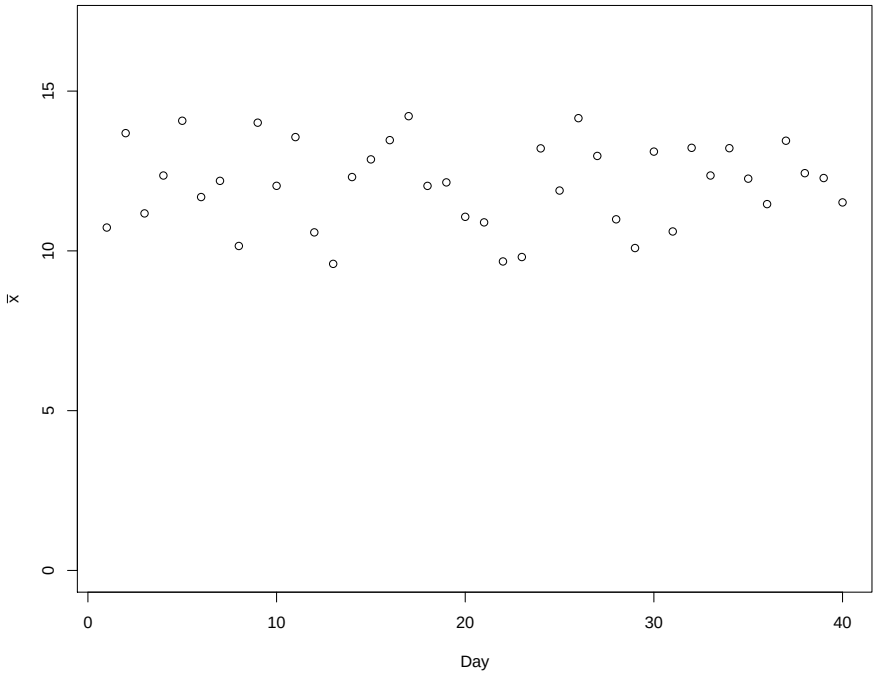

Figure 1: Plot of daily sample means over time, assuming $\mu = 12$, $\sigma = 3$, and a daily sample size $n = 4$.

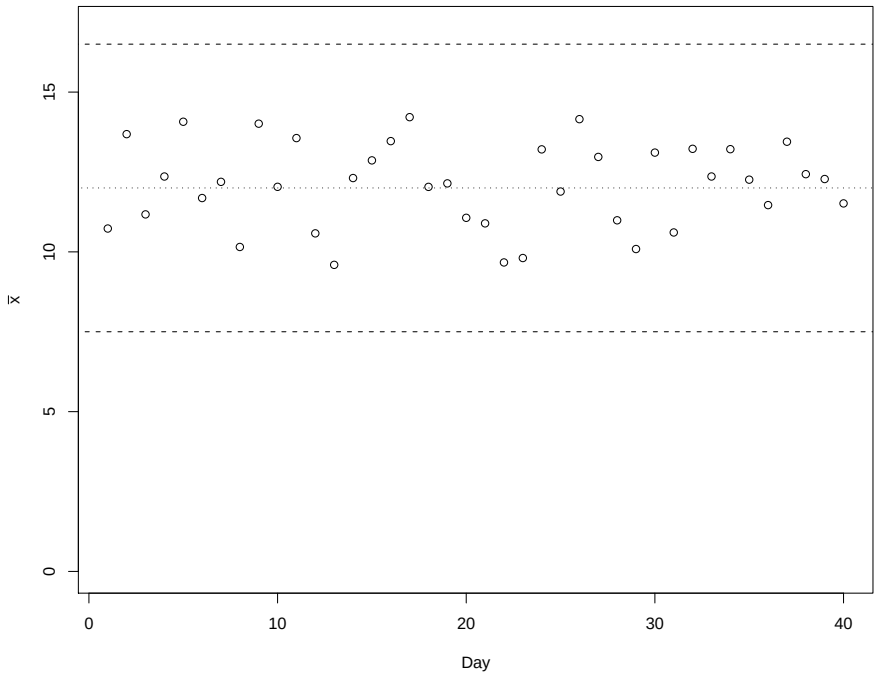

Figure 2: Figure 1, with the addition of the mean line (dotted line), and upper and lower control limits (dashed lines).

mean as a horizontal line across the control chart, and then two more lines, one three times the standard error above the mean ($\mu + \frac{3\sigma}{\sqrt{n}}$) and one three standard errors below the mean ($\mu + \frac{3\sigma}{\sqrt{n}}$). These are the **upper control limit** and **lower control limit**, respectively, abbreviated **UCL** and **LCL**. (See Figure 2.) If the point we plot is within the control limits, all is well; outside, something is wrong. Now we can ask, what's the probability of $\overline{X}$ landing between the LCL and the UCL, if the process is in fact in control?

$$
\begin{aligned}
&\Pr\left(\text{LCL} \leq \overline{X} \leq \text{UCL}\right) = \\
&\quad = \quad \Pr\left(\mu - \frac{3\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + \frac{3\sigma}{\sqrt{n}}\right) \\
&\quad = \quad \Pr\left(-3\frac{\sigma}{\sqrt{n}} \leq Y \leq 3\frac{\sigma}{\sqrt{n}}\right) \text{ where } Y = \overline{X} - \mu \sim \mathcal{N}(0, \sigma/\sqrt{n}) \\
&\quad = \quad \Pr\left(-3 \leq Z \leq 3\right) \text{ where} Z = \frac{\sqrt{n}}{\sigma} Y \sim \mathcal{N}(0,1) \\
&\quad = \quad \Phi(3) - \Phi(-3)
\end{aligned}
$$

$$= \quad 0.9974$$

Let's be very formal about things and state our procedure as an algorithm — something anyone could follow without thinking.

```
Calculate LCL and UCL
Every day
    Calculate X̄
    Is LCL ≤ X̄ ≤ UCL?
        Yes: process is in control; give yourself a donut
        No: process is out of control; sound alarm, find problem
```

There are two ways this procedure can screw up. One is that we sound the alarm even when nothing is really wrong — a **false alarm**. The other is that we don't realize when the process goes out of control — a **miss**. We'll see how to analyze these errors in the next sub-section.

2.1 Reliability Analysis

When evaluating the reliability of our control chart — or any statistical procedure, really — the crucial question is what the **error rates** are — how likely it is to make different kinds of errors. These are also known as the **performance characteristics** or **operating characteristics** of the procedure. We said that there were two kinds of errors we have to guard against here: false alarms, and misses. Let's calculate the probabilities of both kinds of errors.

False alarms first, because they're easier. The probability of a false alarm is called α (you can think of it as standing for "alarm"), and it's the probability that $\overline{X}$ is outside the control limits, even when the process is in control, i.e., even when $\overline{X} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$:

$$\begin{aligned}
\alpha &\equiv \Pr\left((\overline{X} > \mathrm{UCL}) \cup (\overline{X} < \mathrm{LCL})\right) \\
&= 1 - \Pr\left(\mathrm{LCL} \leq \overline{X} \leq \mathrm{UCL}\right) \\
&= 1 - 0.9974 \\
&= 0.0026
\end{aligned}$$

The probability of a miss is called β, because it's not α, and it depends on *how* the process goes out of control. The easiest possibility to consider is that something reaches into the machinery and shifts the mean, but doesn't change the variance, or the form of the distribution. So $X \sim \mathcal{N}(\mu + \delta\sigma, \sigma)$, and $\overline{X} \sim \mathcal{N}(\mu + \delta\sigma, \sigma/\sqrt{n})$. (I'm expressing the shift in the mean in units of the standard deviation, which will make the math easier later.) Now the probability of an error is the probability that we *do* fall within the control limits:

$$\begin{aligned}
\beta(\delta) &= \Pr\left(\mathrm{LCL} \leq \overline{X} \leq \mathrm{UCL}\right) \text{ when } \overline{X} \sim \mathcal{N}(\mu + \delta\sigma, \sigma/\sqrt{n}) \\
&= \Pr\left(\mu - 3\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + 3\frac{\sigma}{\sqrt{n}}\right) \\
&= \Pr\left(-\delta\sigma - 3\frac{\sigma}{\sqrt{n}} \leq Y \leq -\delta\sigma + 3\frac{\sigma}{\sqrt{n}}\right) \text{ where } Y = \overline{X} - \mu - \delta\sigma \sim \mathcal{N}(0, \sigma/\sqrt{n})
\end{aligned}$$

$$
\begin{aligned}
&= \Pr\left(-3-\delta\sqrt{n} \leq Z \leq 3-\delta\sqrt{n}\right) \text{ where } Z = \frac{\sqrt{n}}{\sigma}Z \sim \mathcal{N}(0,1)\\
&= \Phi(3-\delta\sqrt{n}) - \Phi(-3-\delta\sqrt{n})
\end{aligned}
$$

If $\delta\sqrt{n} \gg 3$, then $\beta(\delta) \approx \Phi(-\delta\sqrt{n}) - \Phi(\delta\sqrt{n}) \approx 0$. In other words, if you've either got a big shift in the process (δ is large) or a lot of sample every day ($\sqrt{n}$ is large), the probability of a miss is small. On the other hand, if $\delta\sqrt{n}$ is very small, $\beta \approx \Phi(3) - \Phi(-3) = 1 - \alpha = 0.9974$, i.e., we're very likely to miss a very small shift, because the process still looks like it's *almost* in control.

All of this has been assuming that we set the upper and lower limits of control at the mean $\pm$ 3 times the standard error. There's nothing magic about the number 3 there, though it is conventional. We could use any number κ we like — $\kappa = 1$, or $= 7$, or $= 2.5$, as we like. What happens as we change κ? Repeating the same analysis as before, we'd see that

$$
\begin{aligned}
\alpha(\kappa) &= 1 - \Phi(\kappa) + \Phi(-\kappa)\\
&= 2\left(1 - \Phi(\kappa)\right)\\
\beta(\delta,\kappa) &= \Phi(\kappa - \delta\sqrt{n}) - \Phi(-\kappa - \delta\sqrt{n})
\end{aligned}
$$

As κ grows, α will decrease (because $\Phi(\kappa) \to 1$). But, on the other hand, as κ grows, β will also increase (because $\Phi(\kappa - \delta\sqrt{n}) \to 1$, $\Phi(-\kappa - \delta\sqrt{n}) \to 0$). There is thus an unavoidable trade-off between the two kinds of error — the only way to reduce the risk of false alarms is to increase the risk of misses, and vice versa. The traditional choice of $\kappa = 3$ implicitly chooses to make this trade-off in a certain way; your actual situation might suggest a different trade-off.

Suppose that the process is in control. How long can we expect to go before we get a false alarm? The probability that we get a false alarm on the very first day is α. The probability that we get our first false alarm is $(1-\alpha)\alpha$, because the first day has to have no alarm. In general, the probability that we get our first false alarm on day r is

$$
p(r) = (1-\alpha)^{r-1}\alpha
$$

From chapter 3 and the associated notes, we should remember that this is a **geometric random variable** with probability of success α. So

$$
\mathbf{E}\left[R\right] = \sum_{r=1}^{\infty} r(1-\alpha)^{r-1}\alpha = \alpha\sum_{r=1}^{\infty} r(1-\alpha)^{r-1} = \alpha\left(\frac{1}{1-(1-\alpha)}\right)^2 = \frac{1}{\alpha}
$$

This is known as the **average run length**. On the other hand, if the process is out of control, and the miss rate is β, the number of days we have to go before detecting that it's out of control is again a geometric variable, but now the probability of success is $1-\beta$, and the average run length is $1/(1-\beta)$. If $\beta \ll 1$, then the average run length $\approx 1 + \beta$, which is nice and short; but if β is substantial, we could go for quite some time before noticing that the process is out of control. The next lecture will be about methods which try to shorten the average run length, without messing up the error rates too badly.

3 Beyond the Known Gaussian Case

In everything we've done so far, we've assumed that the process is Gaussian, and that we know the Gaussian parameters μ and σ. This is often not *exactly* the case. Let's consider what to do when we can't make those assumptions.

3.1 Gaussian, Unknown Parameters

If we know X is Gaussian, but don't know its parameters beforehand, we have to estimate them from samples. The easiest way to do this is to take a bunch of samples, say k of them, and compute the mean $\overline{x_i}$ and the sample standard deviation s_i for each one. We then use those to estimate the population mean and the population standard deviation, approximating μ as the mean of means:

$$\overline{\overline{x}} = \frac{1}{k}\sum_{i=1}^{k}\overline{x_i}$$

It'd be nice if we could just do the same thing for the standard deviations, but we can't.

$$\sigma = \frac{1}{k}\sum_{i=1}^{k} s_i \text{ WRONG! WRONG! WRONG!}$$

The problem is that, while the sample mean is, on average, equal to the population mean, this is *not* true of the sample standard deviation.

To anticipate what we're going to see in chapter 6 a little, let's think about estimating some attribute of the population, traditionally called θ, by computing some sample statistic, traditionally called T. The **bias** is the expected difference between T and θ:

$$\text{bias(T)} \equiv \mathbf{E}\left[T - \theta\right] = \mathbf{E}\left[T\right] - \theta$$

If the bias is 0, we say that T is an **unbiased estimator** of θ. The sample mean is an unbiased estimator of μ, but the sample standard deviation can be shown to be a biased estimator of σ. It is then necessary to work out the bias and correct for it.

Through a rather ugly calculation we won't go into, you can show that, if $X \sim \mathcal{N}(\mu, \sigma)$, the sample standard deviation is biased, and, for a sample of size n,

$$\begin{aligned}
\mathbf{E}\left[S\right] &= a_n \sigma \\
\text{where } a_n &= \frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)} \\
\text{and in turn } \Gamma(t) &= \int_0^\infty x^{t-1} e^{-x} dx
\end{aligned}$$

This Γ function is one of the standard "special functions", and there are both pre-calculated tables, and well-developed numerical schemes for approximating it to high precision, so it's not that hard to work out what a_n is for any given

n; for large n, $a_n \to 1$. (The book gives values for some typical small values of n.) What this means is that S/a_n *is* an unbiased estimate of σ. So we multiply all of our sample standard deviations by $\frac{1}{a_n}$, and then we average them:

$$\overline{s} = \frac{1}{k}\sum_{i=1}^{k}\frac{s_i}{a_n}$$

Then the control limits are

$$\begin{aligned} \text{UCL} &= \overline{\overline{x}} + 3\frac{\overline{s}}{\sqrt{n}} \\ \text{LCL} &= \overline{\overline{x}} - 3\frac{\overline{s}}{\sqrt{n}} \end{aligned}$$

and everything proceeds as before.

The book also discusses another way of estimating σ, based on the sample ranges rather than the sample standard deviations. It's pretty much the same idea, only the correction factors are different, and larger. Because it takes fewer steps to compute the range than the standard deviation, people in the before-computer Dark Ages used to prefer the range-based method; there's less point to it these days, however.

3.2 Non-Gaussian Processes

There are three ways we handle non-Gaussian processes.

1. *Pretend they're Gaussian.* We can just act as though they're Gaussians with the appropriate mean and variance. This is actually quasi-legitimate, in the case where the distribution is either binomial or Poisson. The former arises when we're looking at the proportion or fraction of items with some property — say, the fraction defective — and the latter when we're looking at counts per item — dents per car, mis-prints per textbook, etc. Both the binomial and the Poisson distribution are approximately Gaussian, in the right limits, so we can use our Gaussian techniques on them, with a not-too-unclean conscience. Section 16.4 of the textbook goes over this in some detail.

2. *Use the distribution function.* If we know the distribution of $\overline{X}$, we can find an interval — a LCL and UCL — where the probability that LCL $\leq \overline{X} \leq$ UCL is whatever we like, say, 0.9974, or $1 - \alpha$ for any α we choose. In other words, we can chose an interval with the right **coverage**. Then we can proceed as before, knowing that our false alarm rate is just α.

3. *Use* 3σ. There are two reasons to do this. One is that our boss is an engineer named Fred who's been doing this for fifty years, and fifty years ago Fred was told to use 3σ, and if it was good enough for him, it's good enough for us. This is actually not completely unreasonable: if everyone's

control limits are consistent with each other, people will have an easier time communicating and solving problems. Also, even if we can't use the standard Gaussian tables to calculate the error probabilities, we can use Chebyshev's inequality to bound them. Knowing that the control limits are set at 3 times the standard error, we know that $\alpha \leq 1/3^2 = 1/9 \approx 0.1111\ldots$. This is much larger than 0.0026, but it's better than nothing.

Let me say just a little more about option #2, using the distribution function. If X is Gaussian, then $\overline{X}$ is Gaussian, too, and we can work out a control interval with the desired coverage probability analytically. If X is not Gaussian, then $\overline{X}$ tends towards a Gaussian as n grows. (That's what the Central Limit Theorem tells us.) For n small but > 1, the distribution of $\overline{X}$ is generally quite messy, and analytical calculations are hard or impossible. This is why Fred wants you to either assume that everything's Gaussian, or just use 3σ as a rule of thumb. These days, however, we don't need to do analytical calculations. We can *simulate* the sampling, and use that to *approximate* the CDF of the sampling distribution.

For example, Fig. 3 shows the CDF for an exponential distribution with $\lambda = 0.25$. Let's suppose that this is the distribution of X. What's the distribution of $\overline{X} = \frac{1}{3}(X_1 + X_2 + X_3)$? Well, I **could** work it out[4], but I don't want to, and I really don't have to. Fig. 4 shows the CDF I get from simulating a large number of random samples of size three, all drawn from this simulation. The solid line is what I get from simulating 100,000 samples; the dashed line is what I get from only simulating 100 samples. I can now solve numerically for an upper and lower control limits such that $\Pr\left(\text{LCL} \leq \overline{X} \leq \text{UCL}\right) = 0.9974$, and I get UCL = 40.30, LCL = 0 — the upper limit is about 4.1 standard errors above the mean. If instead I used the conventional 3σ formula, I'd get UCL = 32.70, LCL = 0, and — this is the more important bit — the coverage probability would be only 0.9883, which isn't what I really want. (These figures are from using 100,000 samples. If I only used 100 to estimate the CDF, then I can only estimate probabilities to within 1%, and so any UCL between about 36.6 and 40.3 looks equally good to me. But doing the simulation and the calculations with 100,000 values rather than 100 takes at most an extra minute on my laptop.)

4 Control of Process Variability

In addition to controling the main tendency of the process, we will also want to control its variability. In fact, as the book mentions, you will generally want to make sure that the variability is under control, before you start trying to figure out if the main tendency is doing what you want. The principles are almost exactly the same as in controlling the mean — you pick your measure of variability, generally either standard deviation or range, and to try to ensure that it's within acceptable limits. Conventionally, again, the control limits are

[4] It's a matter of taking the convolution of density functions.

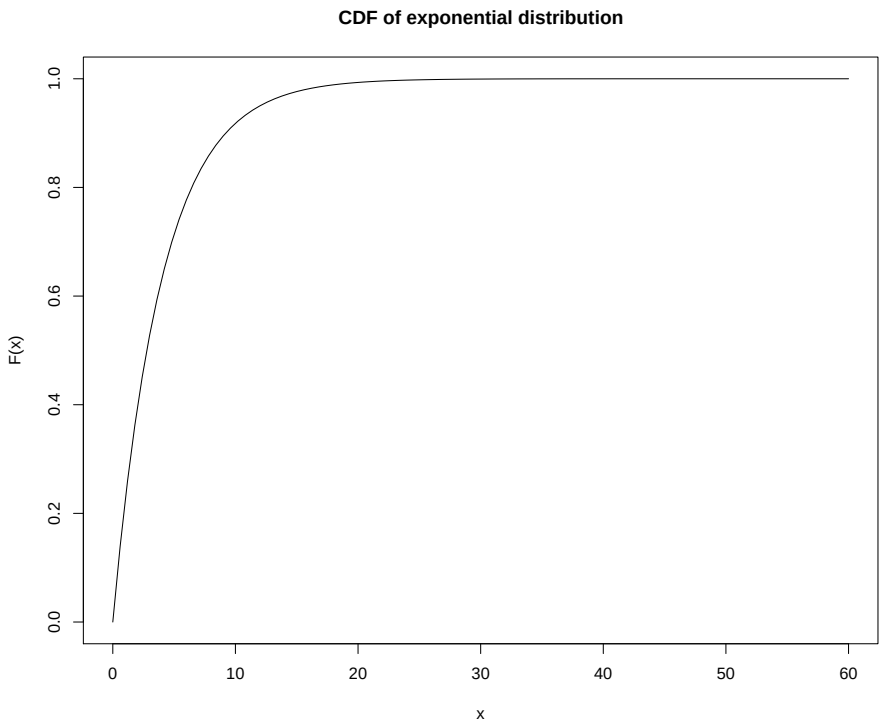

Figure 3: Cummulative distribution function for an exponential distribution, $\lambda = 0.25$.

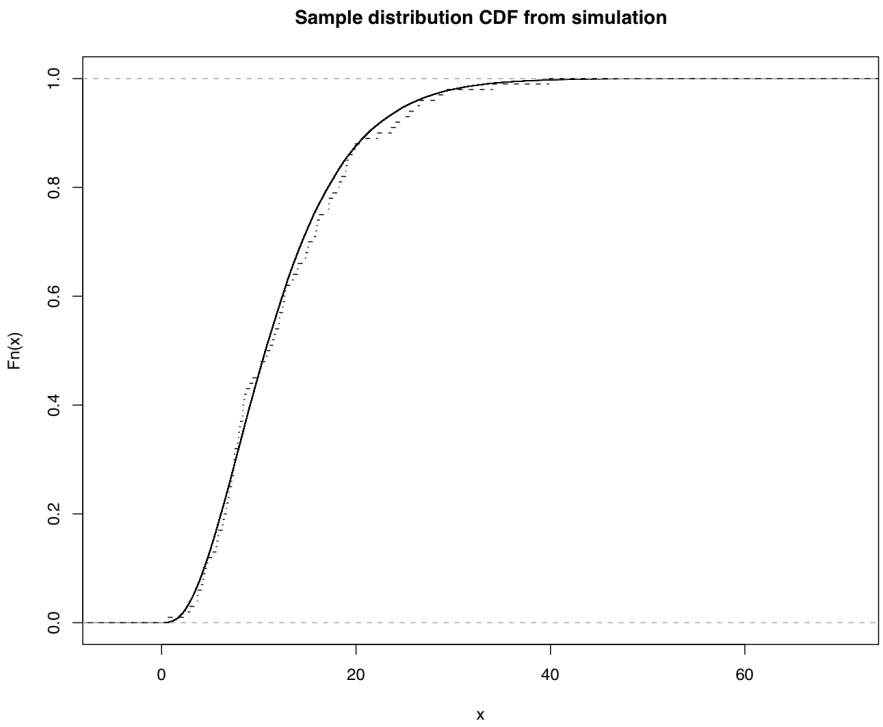

Figure 4: Use of simulation to approximate the CDF for the sample mean, drawing a sample of size $n = 3$ from the exponential distribution in the previous figure. Solid line: 100,000 samples of size 3; dotted line: 100 samples.

given by three times the standard error. Unlike the case of the mean, however, the sample standard deviation and the sample range don't really have Gaussian distributions, even when the underlying process is Gaussian. This makes it much harder to calculate the performance characteristics, unless you're willing to fall back on simulation, as above, or put your trust in tradition, weakly backed by Chebyshev.