Comparison of Systems

CS/ECE 541
1. Stochastic Ordering

Let $X$ and $Y$ be random variables. We say that $X$ is *stochastically larger* than $Y$, denoted $X \succeq_s Y$, if and only if for all $t$,

$$\Pr\{X > t\} \geq \Pr\{Y > t\}$$

An equivalent condition is that there exists a random variable $Y^*$ with the same distribution as $Y$, such that $X \succeq Y^*$.

**Theorem 1.** If $X \succeq_s Y$, then for every monotone non-decreasing function $f$, $E[f(X)] \geq E[f(Y)]$.

The proof follows from the existence of $Y^*$.

A powerful way to compare two stochastic systems is though coupling arguments that establish a stochastic ordering relationship between them.
Example

Let $X$ be exponentially distributed with rate $\lambda_x$, $Y$ be exponentially distributed with rate $\lambda_y$, and have $\lambda_x < \lambda_y$. Then $X \geq_s Y$. For

$$\Pr\{X > t\} = \exp\{-\lambda_x t\}$$
$$> \exp\{-\lambda_y t\}$$
$$= \Pr\{Y > t\}$$

From a coupling point of view, given $X$ we can create $Y^*$ with the distribution of $Y$ such that $X \geq Y^*$. For imagine sampling an instance of $X$ using the inverse CDF method. Sample $u_1$ from a $U[0,1]$ distribution, and define

$$X_1 = -(1/\lambda_x) \log u_1$$

But $-(1/\lambda_x) \log u_1 > -(1/\lambda_y) \log u_1$, so define $Y^*$ as coupled with $X$ through the inverse CDF generation method.
Imagine two G/G/1 queues, $Q_1$ and $Q_2$ with the same inter-arrival distribution, but for service time distributions, $G_{S,1} \geq_s G_{S,2}$.

**Theorem 2.** Under FCFS queueing, the response time distribution for $Q_1$ is stochastically larger than the response time distribution for $Q_2$.

**Proof:** Consider $Q_1$ and $Q_2$ operating in parallel, driven by the same arrival streams. Let $a_1, a_2, a_3, \ldots$ be the times of arrival to these queues. Let $s_{1,i}$ and $s_{2,i}$ be the service time distributions for the $i^{th}$ arrival in $Q_1$ and $Q_2$, respectively. Since $G_{S,1} \geq_s G_{S,2}$, we can sample $s_{2,i}$ in such a way that $s_{1,i} \geq s_{2,i}$, for all $i$.

Let $d_{1,i}$ and $d_{2,i}$ be the departure times of the $i^{th}$ job from $Q_1$ and $Q_2$, respectively. I claim that $d_{1,i} \geq d_{2,i}$ for all $i$. For consider that in the case of $i = 1$

$$d_{1,1} = a_1 + s_{1,1} \geq a_1 + s_{2,1} = d_{2,1}$$

So the claim is true for $i = 1$. If the claim is true for $i = k - 1$, then

$$d_{1,k} = \max\{a_k, d_{1,k-1}\} + s_{1,k} \geq \max\{a_k, d_{2,k-1}\} + s_{1,k} \quad \text{by the induction hypothesis}$$

$$\geq \max\{a_k, d_{2,k-1}\} + s_{2,k} \quad \text{because } s_{1,k} \geq s_{2,k}$$

$$= d_{2,k}$$

The result follows from the observation that the response time of the $i^{th}$ job is $d_{1,i} - a_i$ for $Q_1$, and $d_{2,i} - a_i$ for $Q_2$. 


Variance Reduction Through Anti-thetic Variables

Recall that if $X$ and $Y$ are random variables, then

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$$

and

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y)$$

This implies that if $X$ and $Y$ are positively correlated,

- the variance of their sum is larger than the sum of their variance, and

- the variance of their difference is smaller than the sum of their variances.

So what???

Suppose system 1 has a random metric $X$, under system 2 that variable has a different distribution $Y$, and you want to estimate whether the metric is smaller under system 1 than under system 2.

You could do $N$ independent runs of system 1, $N$ independent runs of system 2, for the $i^{th}$ run in each compute $Z_i = X_i - Y_i$, and use standard techniques to estimate a confidence interval

$$\hat{\mu}_z \pm t_{\alpha/2, N}\hat{\sigma}_Z \frac{1}{N^{1/2}}$$
The benefits of positive correlation

But notice that when the simulation runs of system 1 and system 2 are independent, then

$$\sigma_Z^2 = \text{var}(X) + \text{var}(Y)$$

but if simulation runs of system 1 and system 2 were actively coupled in a way such that you’d expect $X_i$ and $Y_i$ to be positively correlated, say, $Y^*$ has the distribution of $Y$ but is set up to be positively correlated with $X$, then

$$\sigma_Z^2 = \text{var}(X) + \text{var}(Y^*) - 2\text{cov}(X, Y^*) \leq \text{var}(X) + \text{var}(Y)$$

Bottom line: when comparing two systems to determine which is “better”, induced coupling can shrink the confidence interval width for a given number of replications.
Importance Sampling

Another technique for variance reduction is called *importance sampling*

Let $X$ be a random variable with density function $f$ and let $h$ be a function. Then

$$
\mu = E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x) \, dx
$$

We can estimate $E[h(X)]$ by sampling $x_1, x_2, \ldots, x_n$ from $f$ and take

$$
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} h(x_i)
$$

with sample variance

$$
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (h(x_i) - \hat{\mu})^2.
$$

Now consider a distribution $g$ with the property that $g(x) > 0$ whenever $f(x) > 0$. Then an equivalent equation for $\mu$ is

$$
\mu = \int_{-\infty}^{\infty} h(x) \frac{f(x)}{g(x)} g(x) \, dx
= E[h(X)L(X)]
$$

where the last expectation is taken with respect to $g$. $L(x) = f(x)/g(x)$ is called the likelihood ratio.
it this way....when \( g(x_0) \) is large relative to \( f(x_0) \) (skewing towards some feature of interest), we can correct the over-contribution that \( h(x_0) \) has to \( E[h(X)] \) (with the expectation taken with respect to \( f \)) by multiplying it by \( f(x_0)/g(x_0) \). If \( f(x_0) \) is much smaller than \( g(x_0) \), then the contribution of the sampled value \( h(x_0) \) is correspondingly diminished.

You can use this formulation to estimate \( \mu \) by sampling \( y_i \) in accordance to density function \( g \), and take

\[
\hat{\mu}_{is} = \frac{1}{n} \sum_{i=1}^{n} h(y_i)L(y_i)
\]

The intuition here is that we choose \( g \) to bias the sampling of \( y_i \)'s towards regions where \( h(y_i) \) is comparatively large—where the values that most define \( \hat{\mu}_{is} \) live. Fish where the fish are. The factor \( L(y_i) \) corrects for the biasing.

The challenge is to find sampling distributions \( g(x) \) that yield lower variance. The equations above do not ensure anything but the equivalence of two unbiased estimators.
Example: the state of a system $X$ is 1 if failed, and 2 if not failed.

We can in theory choose $g(x)$ that gives no variance to $h(X)L(X)$: Solve

$$h(x) \frac{f(x)}{g(x)} = 1$$

so that every “sample” has value 1! Just take $g(x) = h(x)f(x)$.

Not practical. Why?

To see that importance sampling gives you want you want, notice that

- for $x$ with $h(x) = 1$, then $g(x) = f(x)$ and $h(x)L(x) = h(x) = 1$,

- for $x$ with $h(x) = 2$, then $g(x) = 2f(x)$ and $h(x)L(x) = h(x)/2 = 1$
and

\[ E_g[h(X)L(X)] = \int_x \text{for system failure} h(x)g(x) \, dx + \int_x \text{for system survival} h(x)g(x) \, dx \]

\[ = \int_x \text{for system failure} f(x) \, dx + 2\int_x \text{for system survival} f(x) \, dx \]

\[ = \Pr\{\text{failure}\} \times 1 + 2 \times \Pr\{\text{survival}\} \]