Approximating Distributions

Statistics 110

Summer 2006

Approximating Distributions

As we've seen so far, it can be useful to approximate quantities, such as

- Bounding probabilities by inequalities.
- Bounding moments by functions of other moments.
- Predictions with linear predictors instead of conditional means
- Using Monte Carlo to approximate moments
- Distributions by other distributions (Central Limit Theorem, Binomial by Normal, etc) [well sort have seen this :)]

Example: Suppose that the radius of a circle is a RV $R \sim Gamma(\alpha, \beta)$. What are the mean and variance of the circle's area?

Since the area is given by $A = \pi R^2$, we could answer this question if we knew the distribution of $Y = R^2 = g(R)$.

The inverse function is $g^{-1}(Y) = \sqrt{Y}$. Thus we can get the density by noting

$$\frac{d}{dY} g^{-1}(Y) = \frac{dR}{dY} = \frac{1}{2\sqrt{Y}}$$

So the density of Y is

$$\begin{aligned} f_Y(y) = f_R(g^{-1}(y)) \left| \frac{dR}{dY} \right| &= \frac{\beta^\alpha (\sqrt{y})^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta\sqrt{y}} \frac{1}{2\sqrt{y}} \\ &= \frac{\beta^\alpha y^{\alpha/2-1}}{2\Gamma(\alpha)} e^{-\beta\sqrt{y}}; \qquad y > 0 \end{aligned}$$

This is something we haven't seen before so we can't plug into a known mean and variance formula. Let's try getting the mean and variance by integrating

$$
\begin{aligned}
E[Y^n] &= \int_0^\infty y^n \frac{\beta^\alpha y^{\alpha/2-1}}{2\Gamma(\alpha)} e^{-\beta\sqrt{y}} dy \\
&= \int_0^\infty \frac{\beta^\alpha y^{\alpha/2+n-1}}{2\Gamma(\alpha)} e^{-\beta\sqrt{y}} dy
\end{aligned}
$$

This doesn't look like a particularly nice integral for any n, so we need to try something else.

Idea: Approximate by Taylor series.

General approach:

Suppose that the RV X has $E[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$ and $g(x)$ is a twice continuously differentiable (i.e. $g''(x)$ exists and is continuous). Then

$$g(x) \approx g(\mu) + g'(\mu)(x - \mu)$$

This suggests that

$$E[g(X)] \approx E[g(\mu) + g'(\mu)(X - \mu)] = g(\mu)$$

and

$$\mathrm{Var}(g(X)) \approx \mathrm{Var}(g(\mu) + g'(\mu)(X - \mu)) = (g'(\mu))^2 \mathrm{Var}(X)$$

Note that these approximations can be improved by taking more terms in the Taylor series approximation. For example, if we add one more term,

$$E[g(X)] \approx E\left[g(\mu) + g'(\mu)(X-\mu) + \frac{1}{2}g''(\mu)(X-\mu)^2\right] = g(\mu)+\frac{1}{2}g''(\mu)\sigma^2$$

and

$$\begin{aligned}\mathrm{Var}(g(X)) &\approx \mathrm{Var}(g(\mu) + g'(\mu)(X-\mu) + \frac{1}{2}g''(\mu)(X-\mu)^2) \\ &= (g'(\mu))^2\mathrm{Var}(X) + \frac{1}{4}(g''(\mu))^2(\mathrm{Var}(X^2) - 4\mu^2\sigma^2)\end{aligned}$$

(This approximation for the variance is rarely used as it involves knowing $\mathrm{Var}(X^2)$ or equivalently $E[X^4]$.)

The accuracy of these approximations depends on how close the function is to linear (or quadratic) over the range of the RV that has high probability.

So for the circle example, $A = g(Y) = \pi R^2$. Thus $g'(R) = 2\pi R$ and recall that $E[R] = \frac{\alpha}{\beta}, \mathrm{Var}(R) = \frac{a}{\beta^2}$. So these give

$$E[A] \approx g\left(\frac{\alpha}{\beta}\right) = \pi\frac{\alpha^2}{\beta^2}$$

and

$$\mathrm{Var}(A) \approx \left(g'\left(\frac{\alpha}{\beta}\right)\right)^2 \mathrm{Var}(R) = \left(2\pi\frac{\alpha}{\beta}\right)^2 \frac{\alpha}{\beta^2} = 4\pi^2\frac{\alpha^3}{\beta^4}$$

Some of you may have noticed that I told a fib earlier. In fact, it is possible to calculate $E[Y] = E[R^2]$ and $E[Y^2] = E[R^4]$ using the variable substitution $u = \sqrt{y}$

Thus the true mean and variance are

$$E[A] = \pi\frac{\alpha(\alpha+1)}{\beta^2} = \text{Approximation} + \pi\frac{\alpha}{\beta^2}$$

$$\mathrm{Var}(A) = \pi^2\frac{4\alpha^3+10\alpha^2+6\alpha}{\beta^4} = \text{Approximation} + \pi^2\frac{10\alpha^2+6\alpha}{\beta^4}$$

Suppose that $\alpha = 100, \beta = 10, E[R] = 10, \mathrm{Var}(R) = 1$

	Truth	Approximation	Abs. Error	Rel. Error
$E[A]$	101π	100π	π	0.99%
$\mathrm{Var}(A)$	$410.06\pi^2$	$400\pi^2$	$10.06\pi^2$	2.45%

What happens if we use the second order approximation to $g(\mu) + \frac{1}{2}g''(\mu)\sigma^2$. In this example, $g''(R) = 2\pi$

$$E[A] \approx g\left(\frac{\alpha}{\beta}\right) + \frac{1}{2}g''\left(\frac{\alpha}{\beta}\right)\frac{\alpha}{\beta^2} = \pi\frac{\alpha^2}{\beta^2} + \frac{1}{2}2\pi\frac{\alpha}{\beta^2} = \pi\frac{\alpha(\alpha+1)}{\beta^2}$$

So in this case, it isn't an approximation, but the true value. Actually this has to be the case for this example. Since $g(R)$ is a quadratic function, the 2nd order Taylor series must give the actual function, i.e.

$$x^2 = \mu^2 + 2\mu^2(x-\mu) + (x-\mu)^2$$

For the same reason, the second order approximation to the variance will also give the true value in this case as well.

This idea can be extended to give approximate distributions as well.

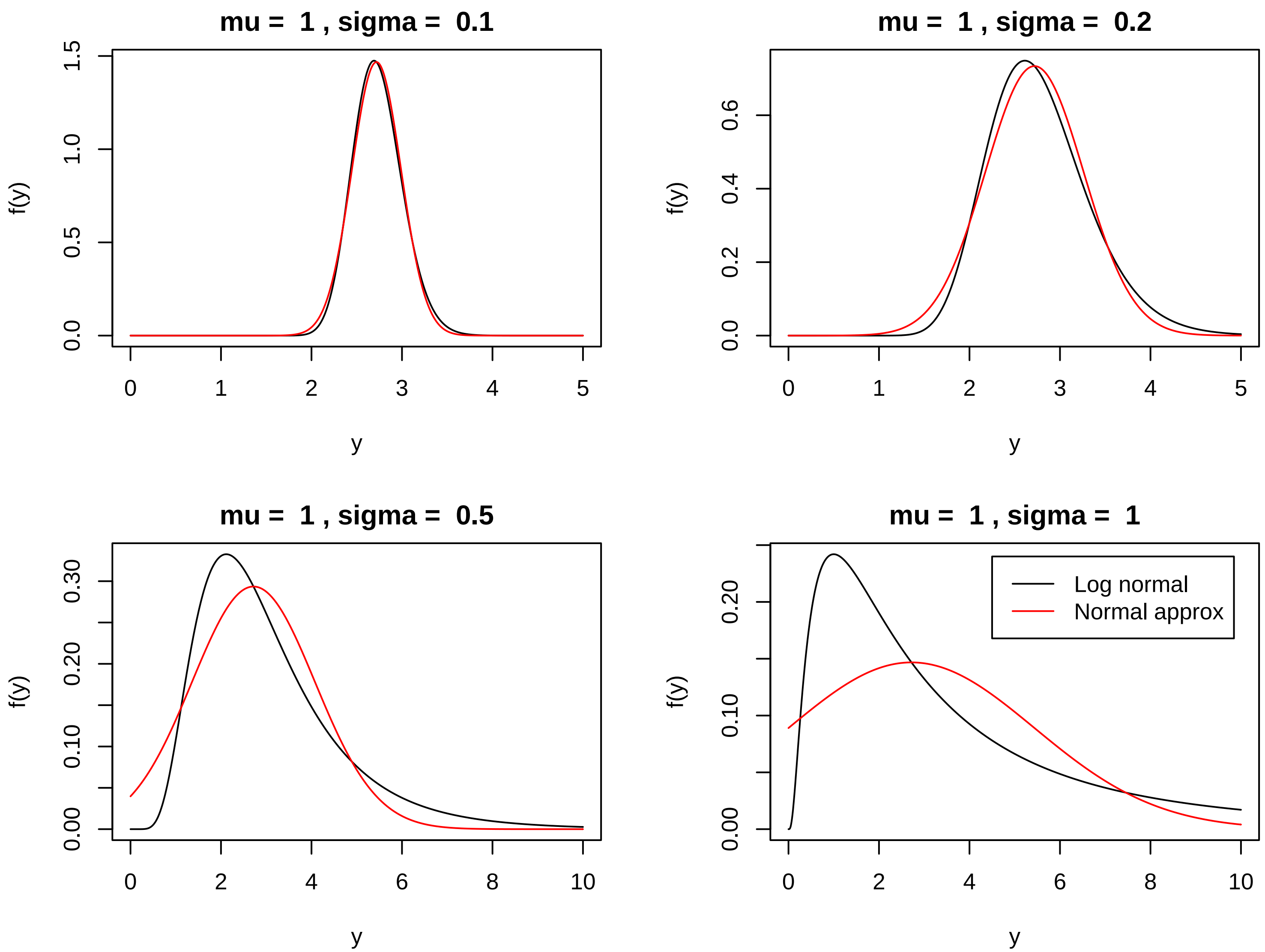
Suppose $X \sim N(\mu, \sigma^2)$ and we are interested in the distribution of $Y = g(X)$ for some function g. Applying the same Taylor series approach gives

$$Y = g(X) \approx g(\mu) + g'(\mu)(X - \mu)$$

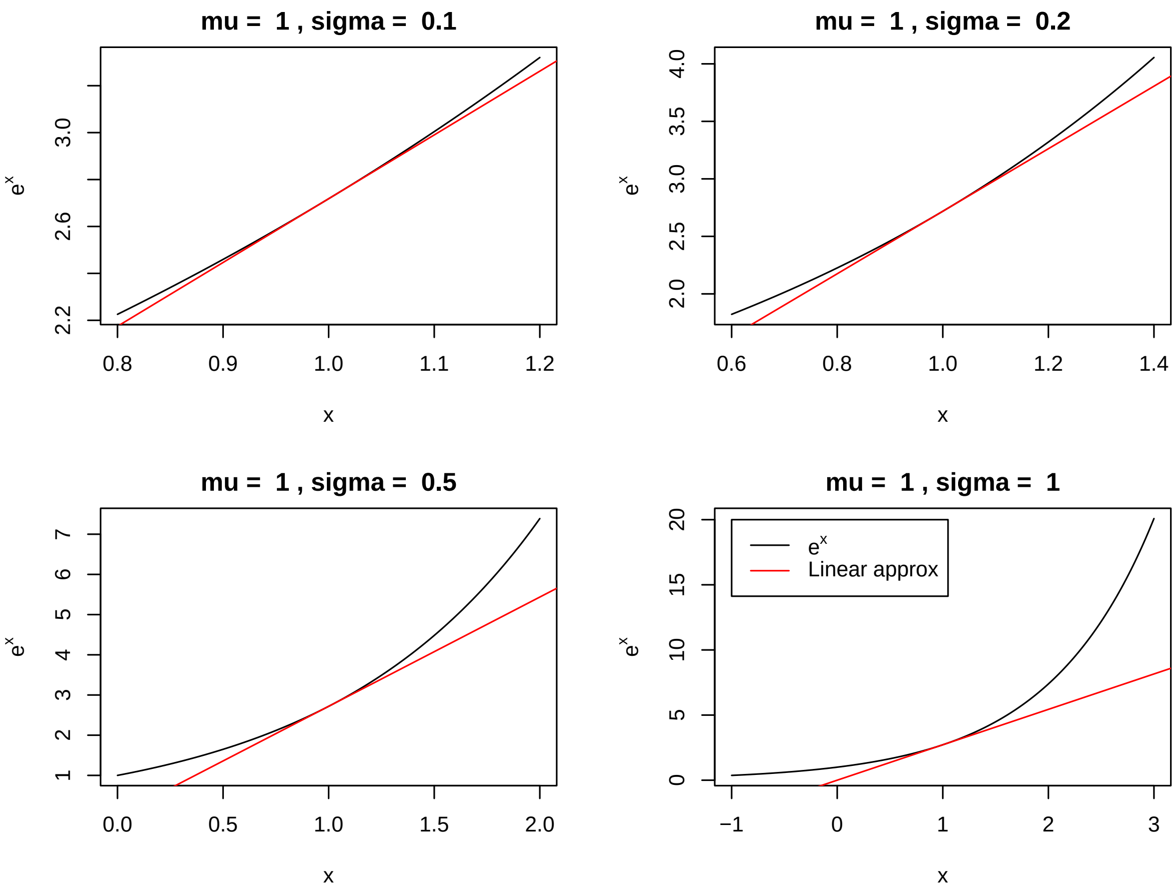
Which suggests Y is approximately normally distributed with mean $g(\mu)$ and variance $(g'(\mu))^2\sigma^2$. (This approach is sometimes known as the delta rule.)

For example, let $Y = e^X$. So $g'(X) = e^X$ so Y is approximately $N(e^\mu, e^{2\mu}\sigma^2)$.

Note that Y actually has a lognormal distribution with $E[Y] = e^{\mu+0.5\sigma^2}$ and $\mathrm{Var}(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$. If σ^2 is small, $e^{\mu+0.5\sigma^2} \approx e^\mu$ and $e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) \approx e^{2\mu}\sigma^2$. However this approximation breaks down as σ increases.


mu = 1 , sigma = 0.1
mu = 1 , sigma = 0.2
mu = 1 , sigma = 0.5
mu = 1 , sigma = 1
f(y)
y
Log normal
Normal approx

As with the moment approximations, the accuracy of this approximation depends on how linear the function over the range of high probability of X.



e^X and the linear approximation over the range $\mu \pm 2\sigma$ which contains 95% of the probability of X.

Note while this example is based on a transformation of a normal RV, the can be used with any distribution. It tends to work best with location-scale distributions and the transformations of normal distributions is the most common use of this technique. Also higher order Taylor series can be used, though the calculations can be ugly.

These ideas can be extended to dealing with functions of multiple RVs. Suppose that $Z = g(X, Y)$, where g is twice differentiable. Then

$$Z = g(X, Y) \approx g(\mu) + \frac{\partial g(\mu)}{\partial x}(X - \mu_X) + \frac{\partial g(\mu)}{\partial y}(Y - \mu_Y)$$

(where $\mu = (\mu_X, \mu_Y)$) which gives

$$E[Z] \approx g(\mu)$$

and

$$\mathrm{Var}(Z) \approx \left(\frac{\partial g(\mu)}{\partial x}\right)^2 \sigma_X^2 + \left(\frac{\partial g(\mu)}{\partial y}\right)^2 \sigma_Y^2 + \left(\frac{\partial g(\mu)}{\partial x}\right)\left(\frac{\partial g(\mu)}{\partial y}\right) \sigma_{XY}$$

The second order Taylor expansion

$$
\begin{aligned}
Z = g(X,Y) \approx g(\mu) &+ \frac{\partial g(\mu)}{\partial x}(X-\mu_X) + \frac{\partial g(\mu)}{\partial y}(Y-\mu_Y) \\
&+ \frac{1}{2}\frac{\partial^2 g(\mu)}{\partial x^2}(X-\mu_X)^2 + \frac{1}{2}\frac{\partial^2 g(\mu)}{\partial y^2}(Y-\mu_Y)^2 \\
&+ \frac{\partial^2 g(\mu)}{\partial x \partial y}(X-\mu_X)(Y-\mu_Y)
\end{aligned}
$$

gives

$$E[Z] \approx g(\mu) + \frac{1}{2}\frac{\partial^2 g(\mu)}{\partial x^2}\sigma_X^2 + \frac{1}{2}\frac{\partial^2 g(\mu)}{\partial y^2}\sigma_y^2 + \frac{\partial^2 g(\mu)}{\partial x \partial y}\sigma_{XY}$$

and something really ugly for $\mathrm{Var}(Z)$.

For example, if $g(X,Y) = \frac{1}{2}XY^2$,

$$\frac{\partial g(\mu)}{\partial x} = \frac{\mu_Y^2}{2}; \qquad \frac{\partial g(\mu)}{\partial y} = \mu_X \mu_Y$$

which gives

$$E[Z] \approx \frac{\mu_X \mu_Y^2}{2}$$

and

$$\mathrm{Var}(Z) \approx \frac{\mu_Y^4}{4}\sigma_X^2 + \mu_X^2 \mu_Y^2 \sigma_Y^2 + \frac{\mu_X \mu_Y^2}{2}\sigma_{XY}$$

The second order approximation to the mean needs

$$\frac{\partial^2 g(\mu)}{\partial x^2} = 0; \qquad \frac{\partial^2 g(\mu)}{\partial y^2} = \mu_X; \qquad \frac{\partial^2 g(\mu)}{\partial x \partial y} = \mu_Y$$

which gives

$$E[Z] \approx \frac{\mu_X \mu_Y^2}{2} + \frac{\mu_X}{2}\sigma_Y^2 + \mu_Y \sigma_{XY}$$

Example (C from Section 4.6): $Z = g(X, Y) = \frac{Y}{X}$

As noted in the text, the necessary partial derivatives are:

$$\frac{\partial g}{\partial x} = \frac{-y}{x^2}; \qquad \frac{\partial g}{\partial x} = \frac{1}{x}$$

$$\frac{\partial^2 g}{\partial x^2} = \frac{2y}{x^3}; \qquad \frac{\partial^2 g}{\partial y^2} = 0; \qquad \frac{\partial^2 g}{\partial x \partial y} = \frac{-1}{x^2}$$

The two approximations to the mean of Z are

- First order:

$$E[Z] = \frac{\mu_y}{\mu_x}$$

- Second order:

$$\begin{aligned} E[Z] &= \frac{\mu_y}{\mu_x} + \sigma_x^2 \frac{\mu_y}{\mu_x^3} - \frac{\sigma_{xy}}{\mu_x^2} \\ &= \frac{\mu_y}{\mu_x} + \frac{1}{\mu_x^2}\left(\sigma_x^2 \frac{\mu_y}{\mu_x} - \sigma_{xy}\right) \end{aligned}$$

Let's assume that $\mu_x \neq 0$, which implies both approximations to $E[Z]$ are well behaved.

Now let's assume that $f_X(0) > 0$ (continuous RV) or $p_X(0) > 0$ (discrete RV). It is possible to show that $E[Z] = \infty$. So these approximations can break down.

However in this setting, if $|\mu_x| >> \sigma_x$, there often is a distribution with mean

$$\begin{aligned} E[Z] &= \frac{\mu_y}{\mu_x} + \sigma_x^2 \frac{\mu_y}{\mu_x^3} - \frac{\sigma_{xy}}{\mu_x^2} \\ &= \frac{\mu_y}{\mu_x} + \frac{1}{\mu_x^2}\left(\sigma_x^2 \frac{\mu_y}{\mu_x} - \sigma_{xy}\right) \end{aligned}$$

that approximates the distribution of Z well.

Note that an approximation to the variance based on the second order Taylor expansion can be made. It will involve up to 4th moments.

Also this idea can be used to get approximate distributions of functions of random variables. Its particularly common when X and Y are bivariate normal. The result of this is that $g(X,Y)$ is approximately normal with the mean and variance given by

$$E[Z] \approx g(\mu)$$

and

$$\mathrm{Var}(Z) \approx \left(\frac{\partial g(\mu)}{\partial x}\right)^2 \sigma_X^2 + \left(\frac{\partial g(\mu)}{\partial y}\right)^2 \sigma_Y^2 + \left(\frac{\partial g(\mu)}{\partial x}\right)\left(\frac{\partial g(\mu)}{\partial y}\right)\sigma_{XY}$$

Note that so far that the Taylor series approximations are done around the mean of X. However they don't need to be done there.

Instead it might make more sense to do it around a different point, such as where $g'(x) = 0$. It depends on what you are trying to do. However around the mean is the usual approach.