Note: Differentiating expectations of a function of a random variable with respect to location and scale parameters

M. Rule

May 16, 2012

Consider a real-valued random variable with a known probability distribution $\Pr(z) = \phi(z)$. From $\phi(z)$, one can generate a scale/location family of probability densities by scaling and shifting $\phi(z)$:

$$\Pr\left(x; \mu, \sigma\right) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$$

The most familiar example of such a family is the univariate Gaussian distribution, when $\phi(z) = [2\pi]^{-1/2}\exp\left(-\frac{1}{2}z^2\right)$. Now, consider the expectation of a function of $\langle f(x)\rangle$ with respect to $\Pr\left(x\right)$.

$$\langle f(x)\rangle = \int_{\mathbb{R}} f(x)\Pr(x)dx = \int_{\mathbb{R}} f(x)\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)dx$$

What are the derivatives of $\langle f(x)\rangle$ with respect to μ and σ^2? The answers are:

$$\begin{aligned}\partial_\mu\langle f(x)\rangle &= \langle f'(x)\rangle \\ \partial_{\sigma^2}\langle f(x)\rangle &= \tfrac{1}{2\sigma^2}\left\langle (x-\mu)f'(x)\right\rangle_x\end{aligned} \tag{1}$$

This question often appears in the special case that x is normally distributed; You'll find derivations elsewhere online given in terms of the cumulative distirbution function of the standard normal distribution.

This note outlines a derivation for any scale/location family using elementary calculus. These derivatives can be obtained by considering how perturbing μ or σ shifts and/or scales the probability density.

For the mean, consider the definition of the derivative:

$$\begin{aligned}\frac{d}{d\mu}\langle f(x)\rangle &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\int_{\mathbb{R}} f(x)\frac{1}{\sigma}\phi\left(\frac{x-\epsilon-\mu}{\sigma}\right)dx - \int_{\mathbb{R}} f(x)\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)dx\right\} \\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\int_{\mathbb{R}} f(x)\frac{1}{\sigma}\phi\left(\frac{x-\epsilon-\mu}{\sigma}\right)dx - \langle f(x)\rangle\right\}\end{aligned} \tag{2}$$

Let $y = x - \epsilon$. Then perform a change of variables ($dy = dx$ and $x = y + \epsilon$):

$$\begin{aligned}
\frac{d}{d\mu}\langle f(x)\rangle &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\int_{\mathbb{R}} f(y+\epsilon)\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)dy - \langle f(x)\rangle\right\} \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\langle f(x+\epsilon)\rangle - \langle f(x)\rangle\right\} \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\langle f(x)+\epsilon f'(x)\rangle - \langle f(x)\rangle\right\} \\
&= \langle f'(x)\rangle
\end{aligned} \tag{3}$$

For the variance, consider the derivative in σ:

$$\frac{d}{d\sigma}\langle f(x)\rangle = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\int_{\mathbb{R}} f(x)\frac{1}{\sigma+\epsilon}\phi\left(\frac{x-\mu}{\sigma+\epsilon}\right)dx - \langle f(x)\rangle\right\} \tag{4}$$

Let $y = \frac{\sigma}{\sigma+\epsilon}(x-\mu)+\mu$. This gives the change of variables

$$\begin{aligned}
dz &= \frac{\sigma+\epsilon}{\sigma}dy \\
x &= \tfrac{\sigma+\epsilon}{\epsilon}(y-\mu)+\mu = y + \frac{\epsilon}{\sigma}(y-\mu)
\end{aligned} \tag{5}$$

Substituting, and simplifying:

$$\begin{aligned}
\frac{d}{d\sigma}\langle f(x)\rangle &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\int_{\mathbb{R}} f(y+\tfrac{\epsilon}{\sigma}(y-\mu))\frac{1}{\sigma+\epsilon}\phi\left(\frac{y-\mu}{\sigma+\epsilon}\right)\frac{\sigma+\epsilon}{\sigma}dy - \langle f(x)\rangle\right\} \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\int_{\mathbb{R}} f(y+\tfrac{\epsilon}{\sigma}(y-\mu))\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma+\epsilon}\right)dy - \langle f(x)\rangle\right\} \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\langle f(y+\tfrac{\epsilon}{\sigma}(y-\mu))\rangle - \langle f(x)\rangle\right\} \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\langle f'(x)\tfrac{\epsilon}{\sigma}(x-\mu)\rangle \\
&= \frac{1}{\sigma}\langle f'(x)(x-\mu)\rangle
\end{aligned} \tag{6}$$

To get the derivative in terms of the variance σ^2, apply the chain rule

$$\frac{d}{d\sigma^2}\langle f(x)\rangle = \frac{d\sigma}{d\sigma^2}\frac{d}{d\sigma}\langle f(x)\rangle = \frac{1}{2\sigma^2}\langle f'(x)(x-\mu)\rangle \tag{7}$$