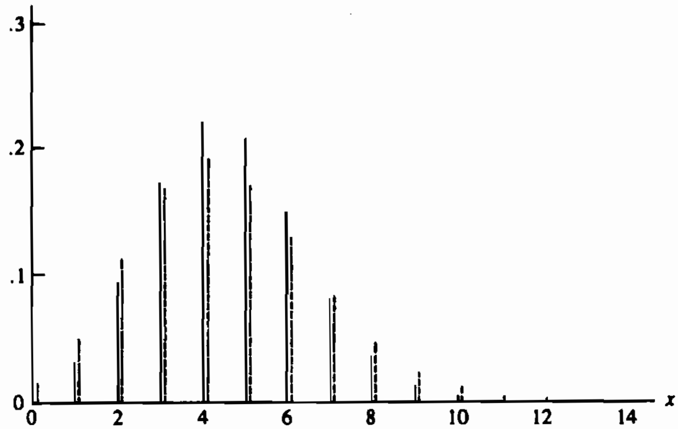

Figure 2.3.3. *Poisson (dotted line) approximation to the binomial (solid line), $n = 15$, $p = .3$*

Proof: By definition,

$$
\begin{aligned}
M_{aX+b}(t) &= \mathrm{E}\left(e^{(aX+b)t}\right) \\
&= \mathrm{E}\left(e^{(aX)t}e^{bt}\right) && \text{(properties of exponentials)} \\
&= e^{bt}\mathrm{E}\left(e^{(at)X}\right) && (e^{bt} \text{ is constant}) \\
&= e^{bt}M_X(at), && \text{(definition of mgf)}
\end{aligned}
$$

proving the theorem. □

2.4 Differentiating Under an Integral Sign

In the previous section we encountered an instance in which we desired to interchange the order of integration and differentiation. This situation is encountered frequently in theoretical statistics. The purpose of this section is to characterize conditions under which this operation is legitimate. We will also discuss interchanging the order of differentiation and summation.

Many of these conditions can be established using standard theorems from calculus and detailed proofs can be found in most calculus textbooks. Thus, detailed proofs will not be presented here.

We first want to establish the method of calculating

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,dx, \tag{2.4.1}$$

where $-\infty < a(\theta), b(\theta) < \infty$ for all θ. The rule for differentiating (2.4.1) is called Leibnitz's Rule and is an application of the Fundamental Theorem of Calculus and the chain rule.

Theorem 2.4.1 (Leibnitz's Rule) *If $f(x,\theta)$, $a(\theta)$, and $b(\theta)$ are differentiable with respect to θ, then*

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,dx = f(b(\theta),\theta)\frac{d}{d\theta}b(\theta) - f(a(\theta),\theta)\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)\,dx.$$

Notice that if $a(\theta)$ and $b(\theta)$ are constant, we have a special case of Leibnitz's Rule:

$$\frac{d}{d\theta}\int_a^b f(x,\theta)\,dx = \int_a^b \frac{\partial}{\partial\theta} f(x,\theta)\,dx.$$

Thus, in general, if we have the integral of a differentiable function over a finite range, differentiation of the integral poses no problem. If the range of integration is infinite, however, problems can arise.

Note that the interchange of derivative and integral in the above equation equates a partial derivative with an ordinary derivative. Formally, this must be the case since the left-hand side is a function of only θ, while the integrand on the right-hand side is a function of both θ and x.

The question of whether interchanging the order of differentiation and integration is justified is really a question of whether limits and integration can be interchanged, since a derivative is a special kind of limit. Recall that if $f(x,\theta)$ is differentiable, then

$$\frac{\partial}{\partial\theta} f(x,\theta) = \lim_{\delta\to 0}\frac{f(x,\theta+\delta)-f(x,\theta)}{\delta},$$

so we have

$$\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta} f(x,\theta)\,dx = \int_{-\infty}^{\infty}\lim_{\delta\to 0}\left[\frac{f(x,\theta+\delta)-f(x,\theta)}{\delta}\right]dx,$$

while

$$\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x,\theta)\,dx = \lim_{\delta\to 0}\int_{-\infty}^{\infty}\left[\frac{f(x,\theta+\delta)-f(x,\theta)}{\delta}\right]dx.$$

Therefore, if we can justify the interchanging of the order of limits and integration, differentiation under the integral sign will be justified. Treatment of this problem in full generality will, unfortunately, necessitate the use of measure theory, a topic that will not be covered in this book. However, the statements and conclusions of some important results can be given. The following theorems are all corollaries of Lebesgue's Dominated Convergence Theorem (see, for example, Rudin 1976).

Theorem 2.4.2 *Suppose the function $h(x,y)$ is continuous at y_0 for each x, and there exists a function $g(x)$ satisfying*

i. $|h(x,y)| \le g(x)$ *for all x and y,*
ii. $\int_{-\infty}^{\infty} g(x)\,dx < \infty$.

Then

$$\lim_{y\to y_0}\int_{-\infty}^{\infty} h(x,y)\,dx = \int_{-\infty}^{\infty}\lim_{y\to y_0} h(x,y)\,dx.$$

The key condition in this theorem is the existence of a dominating function $g(x)$, with a finite integral, which ensures that the integrals cannot be too badly behaved. We can now apply this theorem to the case we are considering by identifying $h(x,y)$ with the difference $(f(x,\theta+\delta)-f(x,\theta))/\delta$.

Theorem 2.4.3 *Suppose $f(x,\theta)$ is differentiable at $\theta=\theta_0$, that is,*

$$\lim_{\delta\to 0}\frac{f(x,\theta_0+\delta)-f(x,\theta_0)}{\delta}=\left.\frac{\partial}{\partial\theta}f(x,\theta)\right|_{\theta=\theta_0}$$

exists for every x, and there exists a function $g(x,\theta_0)$ and a constant $\delta_0>0$ such that

i. $\left|\dfrac{f(x,\theta_0+\delta)-f(x,\theta_0)}{\delta}\right|\le g(x,\theta_0)$, *for all x and $|\delta|\le\delta_0$,*

ii. $\int_{-\infty}^{\infty} g(x,\theta_0)\,dx<\infty$.

Then

$$\left.\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x,\theta)\,dx\right|_{\theta=\theta_0}=\int_{-\infty}^{\infty}\left[\left.\frac{\partial}{\partial\theta}f(x,\theta)\right|_{\theta=\theta_0}\right]dx. \tag{2.4.2}$$

Condition (i) is similar to what is known as a *Lipschitz condition*, a condition that imposes smoothness on a function. Here, condition (i) is effectively bounding the variability in the first derivative; other smoothness constraints might bound this variability by a constant (instead of a function g), or place a bound on the variability of the second derivative of f.

The conclusion of Theorem 2.4.3 is a little cumbersome, but it is important to realize that although we seem to be treating θ as a variable, the statement of the theorem is for one value of θ. That is, for each value θ_0 for which $f(x,\theta)$ is differentiable at θ_0 and satisfies conditions (i) and (ii), the order of integration and differentiation can be interchanged. Often the distinction between θ and θ_0 is not stressed and (2.4.2) is written

$$\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x,\theta)\,dx=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x,\theta)\,dx. \tag{2.4.3}$$

Typically, $f(x,\theta)$ is differentiable at all θ, not at just one value θ_0. In this case, condition (i) of Theorem 2.4.3 can be replaced by another condition that often proves easier to verify. By an application of the mean value theorem, it follows that, for fixed x and θ_0, and $|\delta|\le\delta_0$,

$$\frac{f(x,\theta_0+\delta)-f(x,\theta_0)}{\delta}=\left.\frac{\partial}{\partial\theta}f(x,\theta)\right|_{\theta=\theta_0+\delta^*(x)}$$

for some number $\delta^*(x)$, where $|\delta^*(x)|\le\delta_0$. Therefore, condition (i) will be satisfied if we find a $g(x,\theta)$ that satisfies condition (ii) and

$$\left|\left.\frac{\partial}{\partial\theta}f(x,\theta)\right|_{\theta=\theta'}\right|\le g(x,\theta)\qquad\text{for all } \theta' \text{ such that } |\theta'-\theta|\le\delta_0. \tag{2.4.4}$$

Note that in (2.4.4) δ_0 is implicitly a function of θ, as is the case in Theorem 2.4.3. This is permitted since the theorem is applied to each value of θ individually. From (2.4.4) we get the following corollary.

Corollary 2.4.4 *Suppose $f(x,\theta)$ is differentiable in θ and there exists a function $g(x,\theta)$ such that (2.4.4) is satisfied and $\int_{-\infty}^{\infty} g(x,\theta)\,dx < \infty$. Then (2.4.3) holds.*

Notice that both condition (i) of Theorem 2.4.3 and (2.4.4) impose a uniformity requirement on the functions to be bounded; some type of uniformity is generally needed before derivatives and integrals can be interchanged.

Example 2.4.5 (Interchanging integration and differentiation–I) Let X have the exponential(λ) pdf given by $f(x) = (1/\lambda)e^{-x/\lambda}$, $0 < x < \infty$, and suppose we want to calculate

$$\frac{d}{d\lambda}\mathrm{E}X^n = \frac{d}{d\lambda}\int_0^\infty x^n\left(\frac{1}{\lambda}\right)e^{-x/\lambda}\,dx \tag{2.4.5}$$

for integer $n > 0$. If we could move the differentiation inside the integral, we would have

$$\begin{aligned}\frac{d}{d\lambda}\mathrm{E}X^n &= \int_0^\infty \frac{\partial}{\partial\lambda} x^n\left(\frac{1}{\lambda}\right)e^{-x/\lambda}\,dx \\ &= \int_0^\infty \frac{x^n}{\lambda^2}\left(\frac{x}{\lambda}-1\right)e^{-x/\lambda}\,dx \\ &= \frac{1}{\lambda^2}\mathrm{E}X^{n+1} - \frac{1}{\lambda}\mathrm{E}X^n.\end{aligned} \tag{2.4.6}$$

To justify the interchange of integration and differentiation, we bound the derivative of $x^n(1/\lambda)e^{-x/\lambda}$. Now

$$\left|\frac{\partial}{\partial\lambda}\left(\frac{x^n e^{-x\lambda}}{\lambda}\right)\right| = \frac{x^n e^{-x/\lambda}}{\lambda^2}\left|\frac{x}{\lambda}-1\right| \le \frac{x^n e^{-x/\lambda}}{\lambda^2}\left(\frac{x}{\lambda}+1\right). \qquad \left(\text{since } \tfrac{x}{\lambda} > 0\right)$$

For some constant δ_0 satisfying $0 < \delta_0 < \lambda$, take

$$g(x,\lambda) = \frac{x^n e^{-x/(\lambda+\delta_0)}}{(\lambda-\delta_0)^2}\left(\frac{x}{\lambda-\delta_0}+1\right).$$

We then have

$$\left|\frac{\partial}{\partial\lambda}\left(\frac{x^n e^{-x/\lambda}}{\lambda}\right)\Big|_{\lambda=\lambda'}\right| \le g(x,\lambda) \qquad \text{for all } \lambda' \text{ such that } |\lambda'-\lambda| \le \delta_0.$$

Since the exponential distribution has all of its moments, $\int_{-\infty}^{\infty} g(x,\lambda)\,dx < \infty$ as long as $\lambda - \delta_0 > 0$, so the interchange of integration and differentiation is justified. ||

The property illustrated for the exponential distribution holds for a large class of densities, which will be dealt with in Section 3.4.