Solution

Week 84 (4/19/04)

Poisson and Gaussian

(a) Consider a given box. The probability that exactly x balls end up in it is

$$P(x) = \binom{N}{x}\left(\frac{1}{B}\right)^x\left(1-\frac{1}{B}\right)^{N-x}. \tag{1}$$

This is true because the probability that a certain set of x balls ends up in the given box is $(1/B)^x$, and the probability that the other $N-x$ balls do not end up in the box is $(1-1/B)^{N-x}$, and there are $\binom{N}{x}$ ways to pick this certain set of x balls.

Let us now make approximations to $P(x)$. If N and B are much larger than x, then $N!/(N-x)! \approx N^x$, and $(1-1/B)^x \approx 1$ (we'll be more precise about these approximations below). Therefore,

$$\begin{aligned}
P(x) &= \frac{N!}{(N-x)!x!}\left(\frac{1}{B}\right)^x\left(1-\frac{1}{B}\right)^{N-x} \\
&\approx \frac{N^x}{x!}\left(\frac{1}{B}\right)^x\left(1-\frac{1}{B}\right)^{N} \\
&\approx \frac{1}{x!}\left(\frac{N}{B}\right)^x e^{-N/B} \\
&\equiv \frac{a^x e^{-a}}{x!}.
\end{aligned} \tag{2}$$

This result is called the Poisson distribution. For what x is $P(x)$ maximum? If we set $P(x) = P(x+1)$, we find $x = a-1$. Therefore, we are most likely to obtain $a-1$ or a balls in a box.

We can also consider eq. (2) to be a function of non-integer values of x, by using Stirling's formula, $x! \approx x^x e^{-x}\sqrt{2\pi x}$. This is valid for large x, which is generally the case we will be concerned with. (But note that eq. (2) is valid for small x, too.) Allowing non-integer values of x, the maximum $P(x)$ occurs halfway between $a-1$ and a, that is, at $x = a-1/2$. You can also show this by taking the derivative of eq. (2), with Stirling's expression in place of the $x!$. Furthermore, you can show that $x = a-1/2$ leads to a maximum $P(x)$ value of $P_{\text{max}} \approx 1/\sqrt{2\pi a}$.

In the real world, x can take on only integer values, of course. So it should be the case that the sum of the $P(x)$ probabilities, from $x=0$ to $x=\infty$, equals 1. And indeed,

$$\begin{aligned}
\sum_{x=0}^{\infty} P(x) &= \sum_{x=0}^{\infty}\frac{a^x e^{-a}}{x!} \\
&= e^{-a}\sum_{x=0}^{\infty}\frac{a^x}{x!} \\
&= e^{-a}e^{a} \\
&= 1.
\end{aligned} \tag{3}$$

Let's now be precise about the approximations we made above.

- $N!/(N-x)! \approx N^x$, because

$$\begin{aligned} \frac{N!}{(N-x)!} &= (1)\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{x-1}{N}\right) \\ &\approx N^x\left(1-\frac{x^2}{2N}\right). \end{aligned} \tag{4}$$

  So we need $x \ll \sqrt{N}$ for this approximation to be valid.

- $(1-1/B)^x \approx 1$, because $(1-1/B)^x \approx 1 - x/B$. So we need $x \ll B$ for this approximation to be valid.

- In going from the second to the third line in eq. (2), we used the approximation $(1-1/B)^N \approx e^{-N/B}$. A more accurate statement is $(1-1/B)^N \approx e^{-N/B}e^{-N/2B^2}$, which you can verify by taking the log of both sides. So we need $N \ll B^2$ for this approximation to be valid.

In practice, the basic requirement is $N \ll B^2$. Given this, the $x_{\text{max}} \approx a \equiv N/B$ value that produces the maximum value of $P(x)$ does satisfy both $x_{\text{max}} \ll B$ and $x_{\text{max}} \ll \sqrt{N}$. So the approximations are valid in the region near x_{max}. These are generally the x values we are concerned with, because if x differs much from x_{max}, then $P(x)$ is essentially zero anyway. This will be clear after looking at the Gaussian expression that we'll derive in part (b).

(b) In showing that a Poisson distribution can be approximated by a Gaussian distribution, it will be easier to work with the log of $P(x)$. Let $y \equiv x - a$. The relevant y in this problem will turn out to be small compared to a, so we will eventually expand things in terms of the small quantity y/a. Using Stirling's formula, $x! \approx x^x e^{-x}\sqrt{2\pi x}$ (for large x, as we are assuming here), and also using the expansion $\ln(1+\epsilon) \approx \epsilon - \epsilon^2/2 + \epsilon^3/3 - \cdots$, we have

$$\begin{aligned} \ln P(x) &= \ln\left(\frac{a^x e^{-a}}{x!}\right) \\ &\approx \ln\left(\frac{a^x e^{-a}}{x^x e^{-x}\sqrt{2\pi x}}\right) \\ &= x\ln a - a - x\ln x + x - \ln\sqrt{2\pi x} \\ &= x\ln(a/x) + (x-a) - \ln\sqrt{2\pi x} \\ &= -(a+y)\ln\left(1+\frac{y}{a}\right) + y - \ln\sqrt{2\pi(a+y)} \\ &= -(a+y)\left(\frac{y}{a} - \frac{y^2}{2a^2} + \frac{y^3}{3a^3} + \cdots\right) + y - \ln\sqrt{2\pi(a+y)} \\ &\approx -\frac{y^2}{2a} + \frac{y^3}{6a^2} - \ln\sqrt{2\pi(a+y)}. \end{aligned} \tag{5}$$

The $y^3/6a^2$ term is much smaller than the $y^2/2a$ term (assuming $y \ll a$), so we may ignore it. Also, we may set $2\pi(a+y) \approx 2\pi a$, with negligible multiplicative error. Therefore, exponentiating bother sides of eq. (5) gives

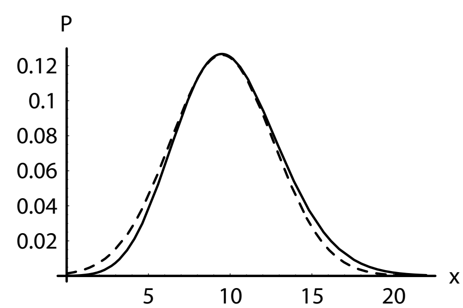
$$P(x) \approx \frac{e^{-(x-a)^2/2a}}{\sqrt{2\pi a}}\,, \tag{6}$$

which is the desired Gaussian distribution. Note that the maximum occurs at $x = a$. If you want to be a little more accuarate, you can include the correction from the $\ln\sqrt{2\pi(a+y)} = \ln\sqrt{2\pi a} + \ln\sqrt{1+(y/a)}$ term in eq. (5), to show that the Gaussian is actually centered at $x = a - 1/2$, that is,

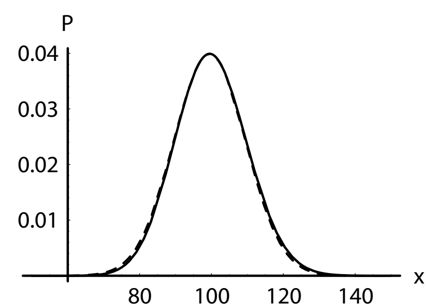
$$P(x) \approx \frac{e^{-\left(x-(a-1/2)\right)^2/2a}}{\sqrt{2\pi a}}\,. \tag{7}$$

The location of the maximum now agrees with the $x_{\max} = a - 1/2$ result that we obtained in part (a). However, for large a, the distinction between a and $a - 1/2$ is fairly irrelevant.

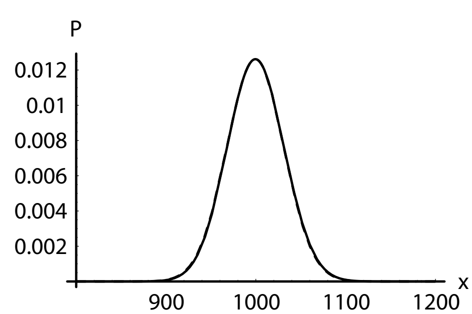
Note that the spread of the Gaussian is of the order $\sqrt{a}$. In our particular case, if we let $x - (a - 1/2) = \eta\sqrt{2a}$, then $P(x)$ is decreased by a factor of $e^{-\eta^2}$ relative to the maximum. $\eta = 1$ gives about 37% of the maximum, $\eta = 2$ gives about 2%, and $\eta = 3$ gives about .01%, which is quite negligible. If, for example, $a = 1000$, then virtually all of the non-negligible part of the graph is contained in the region $900 < x < 1100$, as shown below. The solid curve in these plots is the Poisson distribution from eq. (2), and the dotted curve is the Gaussian distribution from eq. (7). The two curves are essentially indistinguishable in the $a = 1000$ case.



$a = 10$



$a = 100$



$a = 1000$

We should check that the Gaussian probability distribution in eq. (7) has an integral equal to 1. In doing this, we can let the integral run from $-\infty$ to ∞ with negligible error. Using the general result, $\int_{-\infty}^{\infty} e^{-y^2/b}\, dy = \sqrt{\pi b}$, and letting $b \equiv 2a$, we see that the integral is indeed equal to 1.