ПРИЛОЖЕНИЯ

П.1. Распределение Гаусса

Рассмотрим биномиальное распределение, выражаемое формулой (2.14):

$$P(n) = \frac{N!}{n!\,(N-n)!}\,p^n q^{N-n}, \tag{1}$$

где $q=1-p$. Если N велико, то определение вероятности $P(n)$ оказывается затруднительным, так как необходимо вычислять факториалы больших чисел. При этом, однако, оказывается возможным использовать приближения, которые приводят формулу (1) к весьма простому виду.

Мы отмечали еще в п.2.3, что вероятность $P(n)$ имеет максимум, острота которого быстро растет с увеличением N. Это означает, что вероятность $P(n)$ становится пренебрежимо малой, если n заметно отличается от значения $n=\tilde{n}$, соответствующего максимуму. Поэтому значения n, для которых вероятность $P(n)$ *не* является пренебрежимо малой, не отличаются слишком сильно от значения $\tilde{n}$. В этой относительно малой области значений n, обычно и представляющей интерес, легко найти приближенное выражение для $P(n)$.

Таким образом, нам важно исследовать свойства $P(n)$ вблизи значения $\tilde{n}$, отвечающего максимуму. Заметим, что если только p или q не очень близки к 0, это значение $\tilde{n}$ не может лежать близко к 0 или к N. Таким образом, если N велико, то и $\tilde{n}$ является большим числом, и интересующие нас числа n, близкие к $\tilde{n}$, также велики. Но если n велико, то при изменении n на единицу $P(n)$ меняется незначительно, т. е.

$$|P(n+1)-P(n)| \ll P(n),$$

и P является медленно меняющейся функцией n. Поэтому с хорошим приближением P можно считать плавной функцией непрерывной переменной n, хотя физическое значение имеют только целые n. Второе обстоятельство, облегчающее нашу задачу, заключается в том, что логарифм P является значительно медленнее меняющейся

функцией n, чем P. Вместо того чтобы иметь дело непосредственно с P, нам легче исследовать поведение $\ln P$ и найти для этой величины хорошее приближение, пригодное в достаточно большой области изменения переменной n.

Логарифмируя (1), мы имеем

$$\ln P = \ln N! - \ln n! - \ln (N-n)! + n \ln p + (N-n) \ln q. \qquad (2)$$

Значение $n=\tilde{n}$, при котором P достигает максимума, определяется из условия

$$\frac{dP}{dn} = 0$$

или, что эквивалентно, из условия максимума $\ln P$,

$$\frac{d \ln P}{dn} = \frac{1}{P} \frac{dP}{dn} = 0. \qquad (3)$$

Обратим внимание на то, что все факториалы в выражении (2) являются числами, значительно большими единицы. Поэтому мы можем для каждого факториала воспользоваться приближенной формулой (М. 7), из которой следует, что для любого числа m, значительно большего единицы, $m \gg 1$,

$$\frac{d \ln m!}{dm} \approx \ln m. \qquad (4)$$

Таким образом, дифференцируя (2), мы с хорошим приближением имеем

$$\frac{d \ln P}{dn} = -\ln n + \ln (N-n) + \ln p - \ln q. \qquad (5)$$

Чтобы найти максимум P, выражение (5) следует приравнять нулю:

$$\ln \left[\frac{(N-n)}{n} \frac{p}{q} \right] = 0,$$

или

$$\frac{(N-n)}{n} \frac{p}{q} = 1.$$

Поэтому

$$(N-n) p = nq,$$

или

$$Np = n(p+q).$$

Так как $p+q=1$, то значение $n=\tilde{n}$, отвечающее максимуму P, оказывается равным

$$\boxed{\tilde{n} = Np.} \qquad (6)$$

Чтобы исследовать поведение $\ln P$ вблизи максимума, нам следует разложить его в окрестности $\tilde{n}$ в ряд Тейлора. Мы можем

написать

$$\ln P(n) = \\ = \ln P(\tilde{n}) + \left[\frac{d\ln P}{dn}\right] y + \frac{1}{2!}\left[\frac{d^2\ln P}{dn^2}\right] y^2 + \frac{1}{3!}\left[\frac{d^3\ln P}{dn^3}\right] y^3 + \ldots, \quad (7)$$

где

$$y \equiv n - \tilde{n}, \quad (8)$$

а квадратные скобки означают, что производные берутся в точке $n=\tilde{n}$. Так как разложение производится в окрестности точки, соответствующей максимуму P, первая производная [см. (3)] обращается в нуль. Следующие производные могут быть получены последовательным дифференцированием (5). В частности,

$$\frac{d^2\ln P}{dn^2} = -\frac{1}{n} - \frac{1}{N-n} = -\frac{N}{n(N-n)}.$$

Значение этой производной при $n=\tilde{n}$, т. е. при $n=Np$ и $N-n = N(1-p) = Nq$, равно

$$\left[\frac{d^2\ln P}{dn^2}\right] = -\frac{1}{Npq}.$$

Теперь выражение (7) принимает вид

$$\ln P(n) = \ln P(\tilde{n}) - \frac{y^2}{2Npq} + \ldots,$$

или

$$P(n) = \tilde{P} e^{-y^2/2Npq} \ldots = \tilde{P} e^{-(n-\tilde{n})^2/2Npq} \ldots, \quad (9)$$

где мы написали $\tilde{P} \equiv P(\tilde{n})$.

Заметим, что из-за экспоненциального множителя вероятность $P(n)$ в (9) становится пренебрежимо малой по сравнению со значениями в максимуме $\tilde{P}$, если y настолько велико, что $y^2/(Npq) \gg 1$, или $|y| \gg (Npq)^{1/2}$. Вероятность $P(n)$ имеет заметное значение только в области $|y| \leqslant (Npq)^{1/2}$. Здесь величина y достаточно мала для того, чтобы в разложении (7) можно было пренебречь всеми членами высших порядков, кроме оставленного нами члена, пропорционального y^2 *). Мы приходим, таким образом, к выводу, что выражение (9) действительно является хорошим приближением к вероятности $P(n)$ в области, где эта вероятность имеет заметную величину.

Величина $\tilde{P}$ в (9) может быть выражена через p и q с помощью условия нормировки

$$\sum_n P(n) = 1, \quad (10)$$

где суммирование выполняется по всем возможным значениям n. Так как $P(n)$ мало меняется при изменении целого значения на единицу, мы можем заменить сумму (10) интегралом. Область значений n протяженностью dn содержит dn возможных значений $P(n)$.

*) Это верно, пока $(Npq)^{1/2} \gg 1$. См. дополнительную задачу 3.

Таким образом, условие (10) принимает вид

$$\int P(n)\,dn = \int_{-\infty}^{+\infty} \tilde{P} e^{-(n-\tilde{n})^2/2Npq}\,dn = P \int_{-\infty}^{+\infty} e^{-y^2/2Npq}\,dy = 1. \qquad (11)$$

Для облегчения интегрирования мы производим его от $-\infty$ до $+\infty$. Такое приближение является очень хорошим, так как $P(n)$ пренебрежимо мало́ при достаточно больших значениях $|n-\tilde{n}|$. Воспользовавшись формулой (М.23), мы получаем из (11)

$$\tilde{P}\sqrt{2\pi Npq} = 1,$$

или

$$\tilde{P} = \frac{1}{\sqrt{2\pi Npq}}. \qquad (12)$$

Используя этот результат и значение $\tilde{n} = Np$ из (6), мы получаем следующее выражение для вероятности $P(n)$:

$$\boxed{P(n) = \frac{1}{\sqrt{2\pi Npq}} e^{-(n-Np)^2/2Npq}.} \qquad (13)$$

Заметим, что вычисления по формуле (13) гораздо проще вычислений по формуле (1), так как (13) не содержит факториалов.

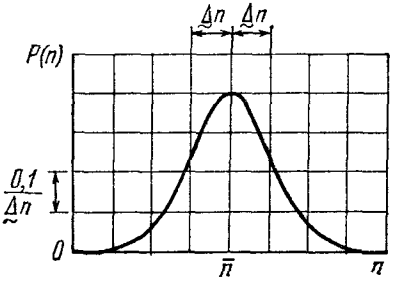

Рис. П.1. Гауссовское распределение плотности вероятности $P(n)$ как функция n. Вероятность $W(x)$ того, что n принимает значения, лежащие между $\tilde{n}-x$ и $\tilde{n}+x$, определяется площадью под кривой, ограниченной этим интервалом. Вычисления показывают, что если Δn означает стандартное отклонение от $\tilde{n}$, то $W(\Delta n)=0{,}683$, $W(2\Delta n)=0{,}954$ и $W(3\Delta n)=0{,}997$.

Вероятность, определяемая формулами (9) и (13), известна как *гауссовское распределение*. Оно было получено нами в результате разложения в ряд логарифма. Поэтому нет ничего удивительного, что гауссовское распределение в статистических рассуждениях появляется очень часто, когда рассматриваемые числа велики.

Выражением (13) для $P(n)$ удобно воспользоваться при вычислении различных средних значений n. Вычисление сумм может быть сведено́ к вычислению эквивалентных интегралов. Это выполняется тем же методом, который был использован при вычислении нормирующего множителя в (11). Мы имеем, таким образом,

$$\bar{n} = \sum P(n)\,dn = (2\pi Npq)^{-1/2} \int_{-\infty}^{+\infty} e^{-(n-Np)^2/2Npq}\, n\,dn =$$

$$= (2\pi Npq)^{-1/2} \int_{-\infty}^{+\infty} e^{-y^2/2Npq} (\tilde{n}+y)\,dy =$$

$$= \tilde{n}(2\pi Npq)^{-1/2} \int_{-\infty}^{+\infty} e^{-y^2/2Npq}\,dy + (2\pi Npq)^{-1/2} \int_{-\infty}^{+\infty} e^{-y^2/2Npq}\, y\,dy.$$

Первый интеграл совпадает с интегралом в (11). Он равен $(2\pi Npq)^{1/2}$. Второй интеграл обращается в нуль; это ясно из соображений симметрии, так как подынтегральное выражение нечетно (оно имеет противоположные знаки для $+y$ и $-y$), и вклады в интеграл при значениях $+y$ и $-y$ уничтожают друг друга. Таким образом, мы получаем

$$\bar{n} = \tilde{n} = Np, \tag{14}$$

т. е. среднее значение n равно значению $\tilde{n} = Np$, при котором вероятность P имеет максимум.

Точно так же мы получаем выражение для дисперсии

$$\overline{(\Delta n)^2} = \overline{(n-\bar{n})^2} = \sum_n P(n)(n-Np)^2 =$$

$$= (2\pi Npq)^{-1/2} \int_{-\infty}^{+\infty} e^{-(n-Np)^2/2Npq} (n-Np)^2\, dn =$$

$$= (2\pi Npq)^{-1/2} \int_{-\infty}^{\infty} e^{-y^2/2Npq}\, y^2\, dy.$$

Используя (М.26), мы получаем, что этот интеграл равен

$$\overline{(\Delta n)^2} = Npq. \tag{15}$$

Таким образом, стандартное отклонение от n равно *)

$$\underset{\sim}{\Delta} n = \sqrt{Npq}. \tag{16}$$

Независимо от того, как мы получили гауссовское распределение, мы можем определить его полностью с помощью двух параметров $\bar{n}$ и $\underset{\sim}{\Delta} n$ [см. (14) и (16)]:

$$P(n) = \frac{1}{\sqrt{2\pi}\, \underset{\sim}{\Delta} n} \exp\left[-\frac{1}{2}\left(\frac{n-\bar{n}}{\underset{\sim}{\Delta} n}\right)^2\right]. \tag{17}$$

Введем переменную

$$z = \frac{n-\bar{n}}{\underset{\sim}{\Delta} n}, \quad \text{так что} \quad n = \bar{n} + (\underset{\sim}{\Delta} n)\, z,$$

с помощью которой выражению (17) можно придать более компактный вид:

$$P(n)\, \underset{\sim}{\Delta} n = \frac{1}{\sqrt{2\pi}} e^{-(1/2)z^2}.$$

Заметим, что гауссовское распределение симметрично относительно переменной z, т. е. значения $P(n)$ для z и $-z$ совпадают.

П.2. Распределение Пуассона

Рассмотрим снова биномиальное распределение (2.14):

$$P(n) = \frac{N!}{n!\,(N-n)!} p^n (1-p)^{N-n}. \tag{18}$$

В предыдущем пункте мы показали, что если $N \gg 1$, то выражение (18) можно приближенно представить гауссовским распределением, и такое приближение справедливо в области значений переменной

*) Заметим, что (14) и (15) точно совпадают с результатами (2.66) и (2.67), полученными в тексте при более общем условии произвольной величины N.