求职中的概率问题

计子毅

November 25,2024

摘要：本文探讨了在找工作时值得思考的概率问题，旨在以找工作为案例，将本学期已经在课堂上学习的概率分布与生活实例进行结合，并且深入理解求职成功率与求职结果趋势，为求职者和即将求职者提供一些想法参考。

在日常生活中，相信各位 byr 都会刷北邮人论坛，通过在论坛上的信息获取，转眼间 2024 年的秋招即将结束了。那么作为求职预备军，值得思考一个问题：求职真的只是将简历投入茫茫的简历大海之中吗？经过细致的观察和思考，可以发现在求职中也能看到一些已知的概率问题，下面就来探讨这些概率问题。

1 正态分布

1.1 正态分布的简要描述

1.1.1 正态分布曲线

正态分布，又称高斯分布。当我们利用高尔顿板进行小球落点实验（如图 1.a），我们会发现，大部分的小球都会集中地落在中间区域。然后进行多次重复实验发现，这种情况并不是偶然，而是一种结果。将小球的分布趋势画成一条连续的曲线，则可以得到一个倒钟型曲线（如图 1.b），这个曲线就是正态分布的曲线。

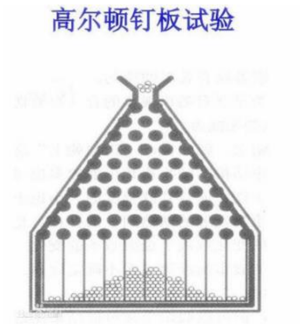

图 1. a

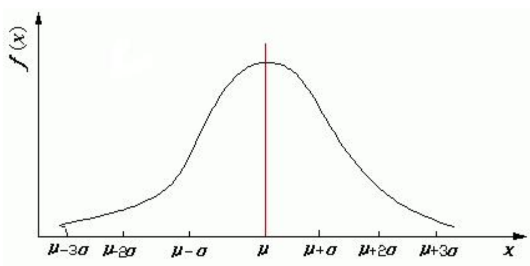

图 1. b

1.1.2 正态分布参数

正态分布是具有两个参数μ和σ²的连续型随机变量的分布，第一参数μ是遵从正态分布的随机变量的均值，第二参数σ²是此随机变量的方差，所以正态分布记作N(μ，σ²)。遵从正态分布的随机变量的概率规律为取μ邻近的值的概率大，而取离μ越远的值的概率越小；σ越小，分布越集中在μ附近，σ越大，分布越分散。正态分布的概率密度函数的特点是：关于μ对称，在μ处达到最大值，在无穷远处取到0，在μ±σ处有拐点。

1.2 个人与行业

1.2.1 在某一行业中所处位置

在某个特定的行业中，就职者的数量n非常大，则n就是一个足够大的样本点集，若将每一个就职人员看成高尔顿钉板试验中的小球，每个人的落点即为这个人在这个行业中的位置。位置越靠右，则表示该就职者在本行业中取得的成就越高；位置越靠左，则表示该就职者在本行业中取得的成就越低。而当样本点足够大时，假设每个人都是随机的取得成就，则最终的结果会如同高尔顿钉板试验，服从正态分布，绝大多数人的落点会集中在对称轴μ附近，这表明绝大多数人在这一行业中是平庸的。

1.2.2 个人在所有行业中的所处位置

当我们把研究主体从某一行业变为某一个求职者，样本空间从就职者变为可选择行业，即此时小球为各个行业，而小球的落点表示该人在这一行业能够取得的成就值。小球A落点越靠右，则该人在行业A中能够取得的成就越高；小球A落点越靠左，则该人在行业A中能够取得的成就越低。根据正态分布的规律，大多数小球会落在中间的区域，这表明这个人在绝大多数行业中既不太差又不太好，也是平庸的。

1.3 个人取得行业顶尖的概率

在前面说到，大多数人在行业中是平庸的，个人在大多数行业中也是平庸的，绝大多数人在绝大多数能力上是平庸的。那就只能承认自己的平庸吗？其实不然，因为描述一个人的特质是如此之多，即使我们在绝大多数领域中都是平均值，但在所有领域都平庸的概率也会小得让人难以置信。俗话说："三百六十行，行行出状元。"，这就是伯努利实验给我们带来的认识。

设在一个行业中处于前 1%的概率假设为 0.01，则至少在一个行业中处于顶尖的概率为 p=1-p(在每个行业都不处于前 1%)=1-（1-0.01）^360=97.32%。则可以得到一个启示：应该尽可能多的尝试新的行业和事物，找到自己喜欢并擅长的东西，才能发挥最大的价值。

2 伯努利试验

在上述正态分布的 1.3 中，我们提及到了伯努利实验，下面就来讨论一下伯努利实验的相关知识点。

2.1 什么是伯努利试验

伯努利试验是一个只有两种结果的简单试验，它的结果是成功或失败，0 或 1，没有中间的立场，则每次试验的结果只有两个，是或不是，或者是多种可能高度抽象成两种情况。

伯努利概型是一种基于独立重复试验的概率模型，它的基本特征在一组固定不变的条件下重复地做一种试验。在伯努利概型中，试验样本总数和概率不能变的情况下，每次试验相同事件发生的概率是一样的；各个重复试验的结果是相互独立互不影响的。且当进行 1 重的伯努利试验时，结果服从 0-1 分布；进行 n 重的伯努利试验，则结果服从二项分布

2.2 伯努利试验相关的三种分布

2.2.1 0-1 分布

只进行 1 次伯努利试验的随机变量，结果只有 2 种，符合 0-1 分布。

2.2.1.(1) 0-1 分布的基本概率和公式

一个随机事件，发生记为 k=1，不发生记为 k=0，若事件服从 0-1 分布，则 k 的分布率如图 2.a 所示。概率分布公式为 f(x)=p^k*(1-p)^(1-k)，其中 k=0 或 1。

X	0	1
p_k	$1-p$	p

图 2.a

2.2.1.(2) 0-1 分布的期望与方差

期望：E(X)=0*(1-p)+1*p=p

则 $E(X^2)=0*(1-p)+1*p=p$

$D(X)=E(X^2)-(E(X))^2=p-p^2=p(1-p)$

2.2.2 几何分布

几何分布就是一种定义为：在 n 次伯努利试验中，试验 k 次才得到第一次成功的机率。详细地说是：前 k-1 次皆失败，第 k 次成功的概率。特点是：每次试验只可能有两种结果。如果只做 1 次试验，那是属于 0-1 分布，但是如果做 N 次试验，但是只有最后一次成功，则随机变量符合 几何分布，但是如果做 N 次试验，没其他限制，则随机变量符合二项分布

由上可知，0-1 分布，几何分布，应该都可以归纳为，二项分布的一种特例。

3 二项分布

在 2.2 中，与伯努利试验相关的三种分布只讨论了两种，因为二项分布与将要讨论的第二个案例相关，在此章结合案例进行讨论。

3.1 二项分布的特征

3.1.1 二项分布的公式

n 次试验中正好得到 k 次成功的概率由概率质量函数给出： p(x=k)=C(n,k)*p^k*(1-p)^(n-k)，其中 k=0，1，2……，其中 n 是总试验次数，k 是成功次数（对应成功的概率 p），C(n,k)= n!/(n-k)!*k!是二项式系数。该公式可以用以下方法理解：我们希望有 k 次成功(p)和 n−k 次失败(1−p)。

3.1.2 二项分布的期望和方差

期望：E(X)=n*p，方差：D(X)=n*p*(1-P)。证明分别如图 3.a,3.b 所示。

概率论中，离散型随机变量期望的定义为

$$E(X) = \sum kP\{X = k\}$$

二项分布概率公式为:

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}$$

则其期望为:

$$E(X) = \sum_{i=0}^{n} iC_n^i p^i (1-p)^{n-i}$$

我们记

$$1 - p = q$$

则

$$E(x) = 0 \times C_n^0 p^0 q^n + 1 \times C_n^1 p^1 q^{n-1} + \cdots + n \times C_n^n p^n q^0$$

因为

$$iC_n^i = i \times \tfrac{n!}{i!(n-i)!} = \tfrac{n!}{(i-1)!(n-i)!} = n \times \tfrac{(n-1)!}{(i-1)!(n-i)!} = nC_{n-1}^{i-1}$$

所以

$$\begin{aligned}
&E(x) = n \times C_{n-1}^{1-1} p^1 q^{n-1} + n \times C_{n-1}^{2-1} p^2 q^{n-2} + \cdots + n \times C_{n-1}^{n-1} p^n q^0 \\
&= np \times C_{n-1}^0 p^0 q^{n-1} + np \times C_{n-1}^1 p^1 q^{n-2} + \cdots + np \times C_{n-1}^{n-1} p^{n-1} q^0 \\
&= np(C_{n-1}^0 p^0 q^{n-1} + C_{n-1}^1 p^1 q^{n-2} + \cdots + C_{n-1}^{n-1} p^{n-1} q^0)
\end{aligned}$$

根据二项式展开定理，有

$$C_{n-1}^0 p^0 q^{n-1} + C_{n-1}^1 p^1 q^{n-2} + \cdots + C_{n-1}^{n-1} p^{n-1} q^0 = (q+p)^{n-1}$$

所以原式

$$\begin{aligned}
&E(x) = np(C_{n-1}^0 p^0 q^{n-1} + C_{n-1}^1 p^1 q^{n-2} + \cdots + C_{n-1}^{n-1} p^{n-1} q^0) \\
&= np(q+p)^{n-1} \\
&= np(1-p+p)^{n-1} \\
&= np
\end{aligned}$$

图 3. a

概率论中，方差的定义为

$$D(X) = E(X^2) - [E(X)]^2$$

因为上文已经得到E(X)，所以现在只需求前者，与上文同理:

$$
\begin{aligned}
&E(X^2) = \sum_{i=0}^{n} i^2 C_n^i p^i q^{n-i} \\
&= C_n^1 p^1 q^{n-1} + \sum_{i=2}^{n} i^2 C_n^i p^i q^{n-i} \\
&= npq^{n-1} + \sum_{i=2}^{n} i n C_{n-1}^{i-1} p^i q^{n-i} \\
&= npq^{n-1} + \sum_{i=2}^{n} (i-1+1) n C_{n-1}^{i-1} p^i q^{n-i} \\
&= npq^{n-1} + \sum_{i=2}^{n} (i-1) n C_{n-1}^{i-1} p^i q^{n-i} + \sum_{i=2}^{n} n C_{n-1}^{i-1} p^i q^{n-i} \\
&= npq^{n-1} + \sum_{i=2}^{n} n(n-1) C_{n-2}^{i-2} p^i q^{n-i} + \sum_{i=2}^{n} n C_{n-1}^{i-1} p^i q^{n-i} \\
&= npq^{n-1} + n(n-1)p^2(q+p)^{n-2} + \sum_{i=1}^{n} n C_{n-1}^{i-1} p^i q^{n-i} - n C_{n-1}^{0} p q^{n-1} \\
&= npq^{n-1} + n(n-1)p^2(q+p)^{n-2} + np(q+p)^{n-1} - npq^{n-1}
\end{aligned}
$$

整理得:

$$
\begin{aligned}
&E(X^2) = n(n-1)p^2(q+p)^{n-2} + np(q+p)^{n-1} \\
&= n(n-1)p^2 + np \\
&= n^2p^2 - np^2 + np \\
&= np(1-p) + n^2p^2 \\
&= npq + n^2p^2
\end{aligned}
$$

综上所述，方差既为:

$$
\begin{aligned}
&D(X) = E(X^2) - [E(X)]^2 \\
&= npq + n^2p^2 - (np)^2 \\
&= npq \\
&= np(1-p)
\end{aligned}
$$

图 3.b

3.2 面试通过概率

在求职中，必将历经一个过程：面试，这个过程往往对于是否能成功入职起到决定性的作用。那么可以提出一个问题：参加面试时，假设两位面试官录取你的概率都是 0.5，且面试采取一票否决制，求面试通过的概率。

那么我们可以设面试官分别为 A 和 B，且二者的决策相互独立，则有随机变量 X，Y，其中

$$X = \begin{cases} 0, & A\text{录取} \\ 1, & A\text{不录取} \end{cases} \quad Y = \begin{cases} 0, & B\text{录取} \\ 1, & B\text{不录取} \end{cases}$$

则 X,Y~B(1,0.5)，设面试不通过的概率为 p，根据对称性：

$$P\{X=0,Y=0\}=P\{X=1,Y=1\}=1-p,$$

$$P\{X=0,Y=1\}=P\{X=1,Y=0\}=p-0.5$$

X，Y 的联合分布率如下：

（X，Y）	0	1
0	1-p	p-0.5
1	p-0.5	1-p

X,Y 联合分布律

因此，0.5≤p≤1，设 X，Y 的相关系数为ρ，则：

$$\rho = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = 4Cov(X,Y)$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=1-p-1/4=3/4-p$$

可得：$p = \frac{3-\rho}{4}$

则有如下情况（注意上述概率 p 为面试不通过的概率）：

① 如果两位面试官的决定总相同，则ρ为 1，此时面试通过的概率为 50%。

② 如果两位面试官的决定总相反，则由一票否决制度，面试通过的概率为 0。

③ 如果两位面试官的决定独立，则相关系数为 0，面试通过的概率为 25%。

3.3 二项分布与正态分布

在 1.1.1 中通过高尔顿钉板试验的结果来表示正态分布的结果，而从直觉上，每个小球在每一层向左或向右下落的概率都是 1/2，那么应该服从一个二项分布 B(n,1/2), n 为钉板的层数。那么事实如此吗？还是说二项分布与正态分布有什么联系？

3.3.1 高尔顿板和二项分布

当 n=1 时，显然小球向左和向右掉落的概率都是 1/2，则服从二项分布 B(n,1/2)。

当 n=N 时，我们假设服从二项分布，则小球落在第 n 层第 k 个格子的概率为：

$$P(N,k) = \frac{C_N^k}{2^N}$$

那么当 n=N+1 时，小球从第 N 层的第 k 格和第 k-1 格都有 1/2 的概率落在第 N+1 层的第 k 格中，即：

$$P(N+1,k)=\frac{1}{2}[P(N,k-1)+P(N,k)]=\frac{1}{2}\left(\frac{C_N^{k-1}}{2^N}+\frac{C_N^k}{2^N}\right)=\frac{C_{N+1}^k}{2^{N+1}}$$

则由数学归纳法可知，高尔顿板的结果服从二项分布。由此引发思考，到底是高尔顿板的结果与正态分布没有关系还是二项分布与正态分布之间存在某种联系？

3.3.2 用二项分布近似正态分布

二项分布是离散的分布，正态分布是连续的分布，两者如何联系起来？先由二项分布在不同 n 下的散点图（图 3.c），我们可以发现，当 n 足够大时，二项分布越来越近似于正态分布。

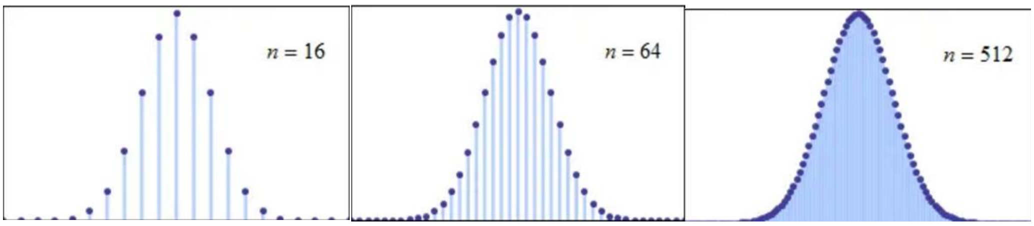

图 3. c

即:

$$\lim_{n\to\infty}P(X_n=k)=\lim_{n\to\infty}C_n^k p^k q^{n-k}=\lim_{n\to\infty}\frac{n!}{k!\,(n-k)!}\cdot p^k q^{n-k},p+q=1$$

由 Stirling 公式$n!\sim\sqrt{2\pi n}(\frac{n}{e})^n$化简得, 并令$\bar{p}=\frac{k}{n}$:

$$\begin{aligned}\lim_{n\to\infty}P(X_n=k)&=\lim_{n\to\infty}\frac{1}{\sqrt{2\pi n\bar{p}(1-\bar{p})}}\left(\frac{p}{\bar{p}}\right)^{n\bar{p}}\left(\frac{1-p}{1-\bar{p}}\right)^{n(1-\bar{p})}\\&=\lim_{n\to\infty}\frac{1}{\sqrt{2\pi n\bar{p}(1-\bar{p})}}e^{-n\left(\bar{p}\ln\left(\frac{\bar{p}}{p}\right)+(1-\bar{p})\ln\left(\frac{1-\bar{p}}{1-p}\right)\right)}\\&=\lim_{n\to\infty}\frac{1}{\sqrt{2\pi n\bar{p}(1-\bar{p})}}e^{-nF(\bar{p})}\end{aligned}$$

将$F(\bar{p})$泰勒展开得:

$$F(\bar{p})\sim\frac{(\bar{p}-p)^2}{2p(1-p)}$$

再代回原式得:

$$\lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} \frac{1}{\sqrt{2\pi n\bar{p}(1-\bar{p})}} e^{-n\frac{(\bar{p}-p)^2}{2p(1-p)}} = \lim_{n\to\infty} \frac{1}{\sqrt{2\pi n\bar{p}(1-\bar{p})}} e^{\frac{(k-np)^2}{2np(1-p)}}$$

再类比正态分布的概率密度公式：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

可知，上式已经具有正态分布的雏形。则当 n 足够大是，可以用二项分布近似的表示正态分布，这是中心极限定理的一种情况，中心极限定理将会在第五章进行学习，不在本文进行讨论。

参考文献：

[1]盛骤，谢式千，潘承毅——《概率论与数理统计》——高等教育出版社

[2]扫 hui——《为什么高尔顿板可以模拟正态分布》——知乎文章

[3]奔跑的犀牛先生——《伯努利试验》——csdn 文章