概率统计模型

- 传送系统的效率
- 报童的诀窍
- 航空公司的预订票策略
- 软件开发人员的薪金
- 教学评估

- **概率模型**

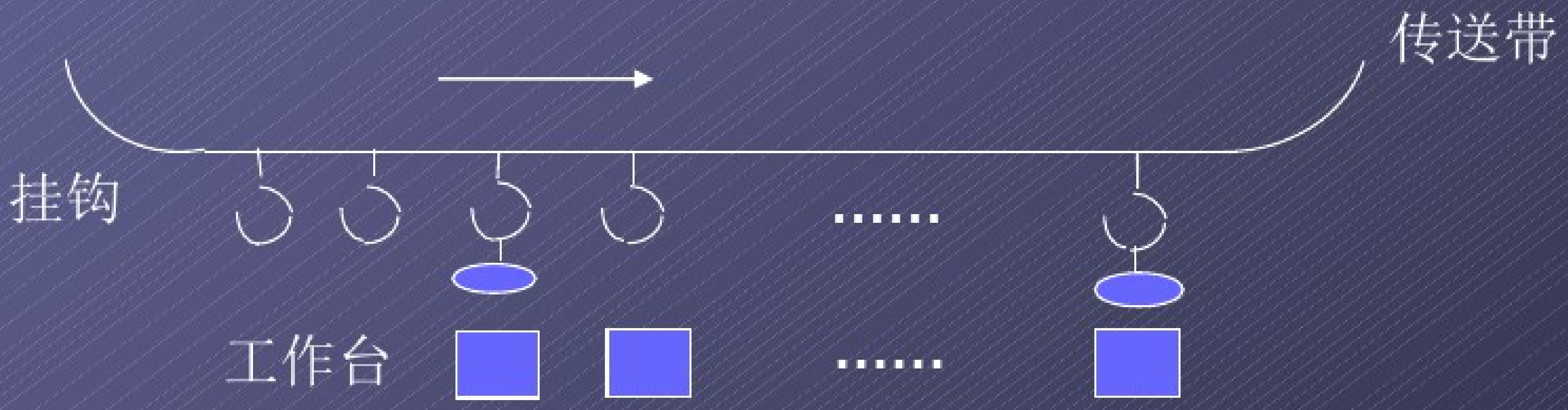
现实世界的变化受着众多因素的影响，包括确定的和随机的。如果从建模的背景、目的和手段看，主要因素是确定的，随机因素可以忽略，或者随机因素的影响可以简单地以平均值的作用出现，那么就能够建立确定性模型。如果随机因素对研究对象的影响必须考虑，就应建立随机模型。本章讨论如何用随即变量和概率分布描述随机因素的影响，建立随机模型--概率模型。

- **统计模型**

如果由于客观事物内部规律的复杂性及人们认识程度的限制，无法分析实际对象内在的因果关系，建立合乎机理规律的模型，那么通常要搜集大量的数据，基于对数据的统计分析建立模型，这就是本章还要讨论的用途非常广泛的一类随机模型—统计回归模型。

一 传送系统的效率

在机械化生产车间里，排列整齐的工作台旁工人们紧张的生产同一种产品，工作台上放一条传送带在运转，带上设置若干钩子，工人将产品挂在经过他上方的钩子上带走，如图。当生产进入稳定状态后，每个工人生产一件产品所需时间是不变的，而他挂产品的时刻是随机的。衡量这种传送系统的效率可以看他能否及时把工人的产品带走。在工人数目不变的情况下传送带速度越快，带上钩子越多，效率越高。



要求构造衡量传送系统效率的指标，并在简化假设下建立模型描述这个指标与工人数目、钩子数量等参数的关系。

1 模型分析

为了用传送带及时带走的产品数量来表示传送系统的效率，在工人生产周期（即生产一件产品的时间）相同的情况下，需要假设工人生产出一件产品后，要么恰好有空钩子经过工作台，他可以将产品挂上带走，要么没有空钩子经过，他将产品放下并立即投入下一件产品的生产，以保证整个系统周期性的运转。

工人生产周期相同，但由于各种因素的影响，经过相当长的时间后，他们生产完一件产品的时刻会不一致，认为是随机的，并在一个生产周期内任一时刻的可能性一样。

由上分析，传送系统长期运转的效率等价于一周期的效率，而一周期的效率可以用它在一周期内能带走的产品数与一周期内生产的全部产品数之比来描述。

2 模型假设

1）有 n 个工人，其生产是独立的，生产周期是常数，n 个工作台均匀排列。

2）生产已进入稳态，即每个工人生产出一件产品的时刻在一个周期内是等可能性的。

3）在一周期内有 m 个钩子通过每一工作台上方，钩子均匀排列，到达第一个工作台上方的钩子都是空的。

4）每个工人在任何时刻都能触到一只钩子，且之能触到一只，在他生产出一件产品的瞬间，如果他能触到的钩子是空的，则可将产品挂上带走；如果非空，则他只能将产品放下。放下的产品就永远退出这个传送系统。

3 模型建立

将传送系统效率定义为一周期内带走的产品数与生产的全部产品数之比，记作 D ，设带走的产品数为 s ,生产的全部产品数为 n ，则 $D=s/n$ 。需求出 s 。

如果从工人的角度考虑，分析每个工人能将自己的产品挂上钩子的概率，这与工人所在的位置有关（如第1个工人一定可挂上），这样使问题复杂化。我们从钩子角度考虑，在稳定状态下钩子没有次序，处于同等地位。若能对一周期内的 m 只钩子求出每只钩子非空的概率 p ，则 $s=mp$ 。

得到 p 的步骤如下：（均对一周期而言）

任一只钩子被一名工人触到的概率是 $1/m$ ；

任一只钩子不被一名工人触到的概率是 $1-1/m$ ；

由工人生产的独立性，任一只钩子不被所有 n 个工人挂上产品的概率，即任一只钩子为空钩的概率是 $\left(1-\frac{1}{m}\right)^n$ ；

任一只钩子非空的概率是 $p=1-\left(1-\frac{1}{m}\right)^n$ 。

传送系统的效率指标为 $D=\frac{mp}{n}=\frac{m}{n}\left[1-\left(1-\frac{1}{m}\right)^{n}\right]$

为了得到比较简单的结果，在钩子数 m 相对于工人数 n 较大，即 $\frac{n}{m}$ 较小的情况下，将多项式 $\left(1-\frac{1}{m}\right)^{n}$ 展开后只取前3项，则有

$$D\approx\frac{m}{n}\left[1-\left(1-\frac{n}{m}+\frac{n(n-1)}{2m^{2}}\right)\right]=1-\frac{n-1}{2m}$$

如果将一周期内未带走的产品数与全部产品数之比记作 E 再假定 $n>>1$，则

$$D=1-E,E\approx\frac{n}{2m}$$

当 $n=10,m=40$ 时,上式给出的结果为 $D=87.5\%$
用 D 的精确表达式计算得 $D=89.4\%$

4 模型评价

这个模型是在理想情况下得到的，其中一些假设，如生产周期不变，挂不上钩子的产品退出系统等是不现实的，但模型的意义在于，一方面利用基本合理的假设将问题简化到能够建模的程度，并用简单的方法得到结果；另一方面所得到的简化结果具有非常简单的意义：指标 $E=1-D$ 与 n 成正比，与 m 成反比。通常工人数目 n 是固定的，一周期内通过的钩子数 m 增加一倍，可使“效率” E 降低一倍。（可理解为相反意义的效率）

思考：

如何改进模型使“效率”降低?

考虑通过增加钩子数来使效率降低的方法：

在原来放置一只钩子处放置的两只钩子成为一个钩对。一周期内通过 m 个钩对，任一钩对被任意工人触到的概率 $p=1/m$，不被触到的概率 $q=1-p$，于是任一钩对为空的概率是 q^n，钩对上只挂一件产品的概率是 npq^{n-1}，一周期内通过的 $2m$ 个钩子中，空钩的平均数是 $m\left(2q^n+npq^{n-1}\right)$

带走产品的平均数是 $2m-m\left(2q^n+npq^{n-1}\right)$

未带走产品的平均数是 $n-\left[2m-m\left(2q^n+npq^{n-1}\right)\right]$

按照上一模型的定义，有

$$E=1-D=1-\frac{m}{n}\left[2-2\left(1-\frac{1}{m}\right)^n-\frac{n}{m}\left(1-\frac{1}{m}\right)^{n-1}\right]$$

利用 $\left(1-\frac{1}{m}\right)^n$ 和 $\left(1-\frac{1}{m}\right)^{n-1}$ 的近似展开，可得

$$E \approx \frac{(n-1)(n-2)}{6m^2} \approx \frac{n^2}{6m^2}$$

注意：$\left(1-\frac{1}{m}\right)^n$ 展开取4项，$\left(1-\frac{1}{m}\right)^{n-1}$ 展开取3项。而上一模型中的方法有 $E_1 = \frac{n}{4m}$ 有 $E = \beta E_1$ $\quad \beta = \frac{2n}{3m}$

当 $m > \frac{2n}{3}$ 时，$\beta < 1$，所以该模型提供的方法比上一个模型好。

二 报童的诀窍

问题：

报童每天清晨从报社购进报纸零售，晚上将没有卖掉的报纸退回。设报纸每份的购进价为b，零售价为a，退回价为c，假设$a>b>c$。即报童售出一份报纸赚$a-b$，退回一份赔$b-c$。报童每天购进报纸太多，卖不完会赔钱；购进太少，不够卖会少挣钱。试为报童筹划一下每天购进报纸的数量，以获得最大收入。

模型分析：

购进量由需求量确定，需求量是随机的。假定报童已通过自己的经验或其他渠道掌握了需求量的随机规律，即在他的销受范围内每天报纸的需求量为 r 份的概率是 $f(r)\ (r=0,1,2,\Lambda)$

有了 $f(r)$ 和 a,b,c 就可以建立关于购进量的优化模型。

模型建立：

假设每天购进量是 n 份，需求量 r 是随机的，r 可以小于，等于或大于 n，所以报童每天的收入也是随机的。那么，作为优化模型的目标函数，不能取每天的收入，而取长期卖报（月，年）的日平均收入。从概率论大数定律的观点看，这相当于报童每天收入的期望值，简称平均收入。

记报童每天购进 n 份报纸的平均收入为 $G(n)$，如果这天的需求量 $r \le n$，则售出 r 份，退回 $n-r$ 份；如果需求量 $r > n$ 则 n 份将全部售出。需求量为 r 的概率是 $f(r)$，则

$$G(n)=\sum_{r=0}^{n}[(a-b)r-(b-c)(n-r)]f(r)+\sum_{r=n+1}^{\infty}(a-b)nf(r)$$

问题归结为在 $f(r),a,b,c$ 已知时，求 n 使 $G(n)$ 最大。

模型求解： 和购进量 n 都相当大，将 r 视为连续变量便于分析和计算，这时概率 $f(r)$ 转化为概率密度函数 $p(r)$

通常需求量

则

r

$$G(n)=\int_0^n[(a-b)r-(b-c)(n-r)]p(r)dr+\int_n^{\infty}(a-b)np(r)dr$$

计算 $\frac{dG}{dn}=(a-b)np(n)-\int_0^n(b-c)p(r)dr$

$$-(a-b)np(n)+\int_n^{\infty}(a-b)p(r)dr$$

$$=-(b-c)\int_0^n p(r)dr+(a-b)\int_n^{\infty}p(r)dr$$

令 $\frac{dG}{dn}=0$，得到 $$\frac{\int_0^n p(r)dr}{\int_n^\infty p(r)dr}=\frac{a-b}{b-c}$$

使报童日平均收入达到最大的购进量 n 应满足上式。

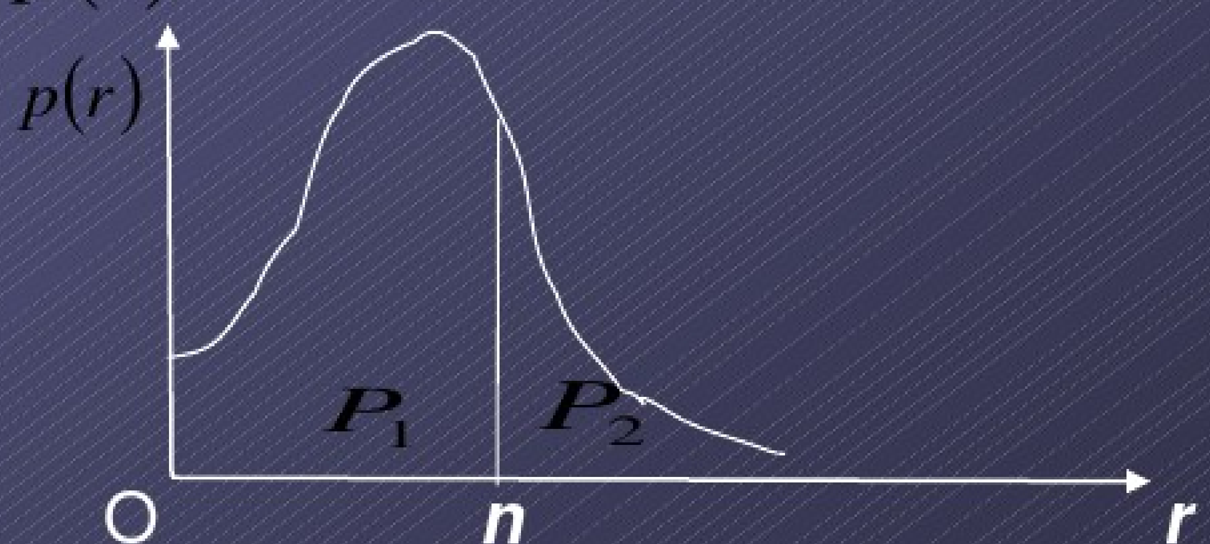
因为 $\int_0^\infty p(r)dr=1$，所以

$$\int_0^n p(r)dr=\frac{a-b}{a-c}$$

根据需求量的概率密度 $p(r)$ 的图形可以确定购进量 n

在图中用 P_1,P_2 分别表示曲线 $p(r)$ 下的两块面积，则

$$\frac{P_1}{P_2}=\frac{a-b}{b-c}$$



因为当购进 n 份报纸时，$P_1=\int_0^n p(r)dr$ 是需求量 r 不超过 n 的概率，即卖不完的概率；$P_2=\int_n^\infty p(r)dr$ 是需求量 r 超过 n 的概率，即卖完的概率，所以上式表明，购进的份数 n 应该使卖不完与卖完的概率之比，恰好等于卖出一份赚的钱 $a-b$ 与退回一份赔的钱 $b-c$ 之比。

结论：

当报童与报社签订的合同使报童每份赚钱与赔钱之比约大时，报童购进的份数就应该越多。

练习：

利用上述模型计算，若每份报纸的购进价为0.75元，售出价为1元，退回价为0.6元，需求量服从均值500份，均方差50份的正态分布，报童每天应购进多少份报纸才能使平均收入最高，最高收入是多少？