- So, $N_2 \sim$ Poisson(11), $P(N_2=k) = 11^k \cdot e^{-11}/k!$ and (2 months)

$$P(N_2 \geq 3) = 1 - P(N_2 \leq 2) = 1 - \sum_{k=0}^{2} \frac{e^{-11} \cdot 11^k}{k!}$$

➢ Summary for $X \sim$ Poisson(λ)

Q: What's the original sample space Ω for the X? For $\omega \in \Omega$, $N_t(\omega)$: a nondecreasing step function of t.

cf.
- the Ω for binomial X (LNp.5-20~21)
- the Ω for negative binomial X (LNp.5-26~27)

- Range: $\mathcal{X} = \{0, 1, 2, \ldots\}$
- Pmf: $f_X(x) = \lambda^x e^{-\lambda}/x!$, for $x \in \mathcal{X}$
- Parameter: $0<\lambda<\infty$
- Mean: $E(X)=\lambda$
- Variance: $Var(X)=\lambda$

10/31

- Hypergeometric Distribution

超幾何

Recall. LNp.4-5~4-7

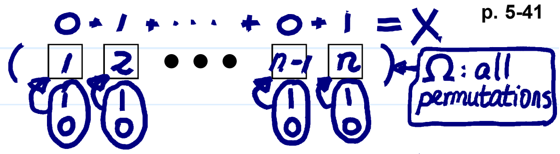
➢ Experiment: Draw a sample of n ($\leq N$) balls *without replacement* from a box containing R red balls and $N-R$ white balls (the balls are equally likely to be drawn)

- Let X be the number of red balls in the sample
- **Q**: What is $P(X=k)$? ← distribution of X is ?
- Example. The Committee Example (LNp.5-6). binomial$(n, \frac{R}{N})$ (check LNp.5-21)
- (cf.) If drawn *with replacement*, what is the distribution of X?

graphs in LNp 5-21

cf.

➢ Probability Mass Function



- Theorem. For $k = 0, 1, 2, \ldots, n$, (possible values of X)

original sample space Ω of X: Ω= {combination of n} $\#\Omega=\binom{N}{n}$

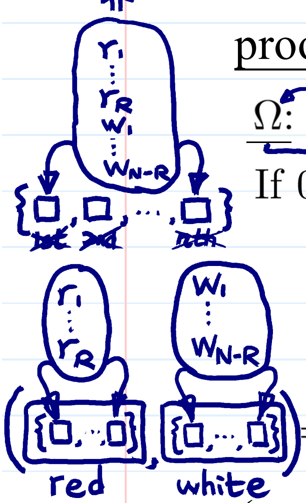
$$P(X=k) = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}.$$

intuition

with replacement ← indep. Bernoulli → $X \sim$ binomial

without replacement ← dep. Bernoulli (△) → $X \sim$ hypergeometric ← cf. graphs in LNp.5-20~21

(Notice that $\binom{r}{t} \equiv 0$ when either $t<0$ or $r<t$.) $\max(0, n+R-N) \leq k \leq \min(R, n)$



proof. Label the N balls as $r_1, \ldots, r_R, w_1, \ldots, w_{N-R}$.

alternative: all $(N)_n$ permutations

Ω: combinations of size n from N different balls. $\Rightarrow \#\Omega = \binom{N}{n}$

→ symmetric outcomes

If $0 \leq k \leq R$ and $0 \leq n-k \leq N-R$,

k red balls may be chosen in $\binom{R}{k}$ ways.

$n-k$ white balls may be chosen in $\binom{N-R}{n-k}$ ways.

$\Longrightarrow \#\{X=k\} = \binom{R}{k}\binom{N-R}{n-k}$

- (exercise) Show that the following function is a pmf.

For (iii) in LNp.5-6, apply △ in LNp.5-42

$$f(k) = \begin{cases} \binom{R}{k}\binom{N-R}{n-k}/\binom{N}{n}, & k = 0, 1, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

(※)

- The distribution of the r.v. X is called the *hypergeometric* distribution with parameters n, N, and R.
  - The hypergeometric distribution is called after the hypergeometric identity:

$$\boxed{i+j=r} \quad (1+x)^a(1+x)^b = \left[\sum_{i=0}^{a}\binom{a}{i}x^i\right]\left[\sum_{j=0}^{b}\binom{b}{j}x^j\right]$$

$$(1+x)^{a+b} = \sum_{r=0}^{a+b}\binom{a+b}{r}x^r$$

$$\boxed{3} - \binom{a+b}{r} = \sum_{k=0}^{r}\binom{a}{k}\binom{b}{r-k}.$$

(why? ∵ (Δ) in LNp.5-41)

➢ Theorem. The mean and variance of hypergeometric(n, N, R) are

$$\boxed{\text{intuitive interpretation}} \quad \mu = \frac{nR}{N} \quad \text{and} \quad \sigma^2 = \frac{nR(N-R)(N-n)}{N^2(N-1)} = n\left(\frac{R}{N}\right)\left(1-\frac{R}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$\frac{N-n}{N-1} = 1-\frac{n-1}{N-1}$

same as with replacement (LNp.5-24) → with replacement, $\sigma^2 = n\left(\frac{R}{N}\right)\left(1-\frac{R}{N}\right)$ cf.

binomial(n, $p=R/N$)

proof.

$$E(X) = \sum_{x=0}^{n} x \cdot \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} = \sum_{x=1}^{n} \cancel{x} \cdot \frac{R!}{\cancel{x}!(R-x)!} \cdot \frac{\binom{N-R}{n-x}}{\binom{N}{n}}$$

$R-1$, $x-1$, $N-1$, $n-1$

$$= \frac{nR}{N}\sum_{x=1}^{n}\frac{\binom{R-1}{x-1}\binom{(N-1)-(R-1)}{(n-1)-(x-1)}}{\binom{N-1}{n-1}} \overset{\text{Let } y=x-1}{=} \frac{nR}{N}\sum_{y=0}^{n-1}\underbrace{\frac{\binom{R-1}{y}\binom{(N-1)-(R-1)}{(n-1)-y}}{\binom{N-1}{n-1}}}_{\text{pmf of hypergeometric}(n-1,N-1,R-1)} = \frac{nR}{N}$$

$$E[X(X-1)] = E(X^2-X) = E(X^2)-E(X)$$

$$= \sum_{x=0}^{n} x(x-1)\cdot\frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} = \sum_{x=2}^{n}\cancel{x(x-1)}\cdot\frac{R!}{\cancel{x}!(R-x)!}\cdot\frac{\binom{N-R}{n-x}}{\binom{N}{n}}$$

$R-2$, $x-2$, $N-2$, $n-2$

$$= \frac{n(n-1)R(R-1)}{N(N-1)}\sum_{x=2}^{n}\frac{\binom{R-2}{x-2}\binom{(N-2)-(R-2)}{(n-2)-(x-2)}}{\binom{N-2}{n-2}}$$

Let $y=x-2$

$$\overset{}{=} \frac{n(n-1)R(R-1)}{N(N-1)}\sum_{y=0}^{n-2}\underbrace{\frac{\binom{R-2}{y}\binom{(N-2)-(R-2)}{(n-2)-y}}{\binom{N-2}{n-2}}}_{\text{pmf of hypergeometric}(n-2,N-2,R-2)} = \frac{n(n-1)R(R-1)}{N(N-1)}$$

$$Var(X) = E(X^2)-[E(X)]^2 = [E(X^2)-E(X)]+E(X)-[E(X)]^2$$

$$= \frac{n(n-1)R(R-1)}{N(N-1)} + \frac{nR}{N} - \left(\frac{nR}{N}\right)^2 = \frac{nR(N-R)(N-n)}{N^2(N-1)}$$

Recall. binomial ↕ Poisson

c.f.

➢ Theorem. Let $N_i \to \infty$ and $R_i \to \infty$ in such a way that

$$p_i \equiv R_i/N_i \to p, \quad n \text{ fixed}$$

where $0 < p < 1$, then

Intuition: When # of red & white balls are very large, n relatively small, without replacement ≈ with replacement

$\boxed{n \ll N_i}$

pmf of hypergeometic (dep. case) →

$$\frac{\binom{R_i}{k}\binom{N_i-R_i}{n-k}}{\binom{N_i}{n}} \to \binom{n}{k}p^k(1-p)^{n-k}.$$

← pmf of binomial (indep. case)

proof.

$$\frac{\binom{R_i}{k}\binom{N_i-R_i}{n-k}}{\binom{N_i}{n}} = \frac{R_i!}{k!(R_i-k)!}\cdot\frac{(N_i-R_i)!}{(n-k)![(N_i-R_i)-(n-k)]!}\cdot\frac{n!(N_i-n)!}{N_i!}$$

$$= \frac{n!}{k!(n-k)!}\cdot\left[\underbrace{\frac{R_i}{N_i}}_{\to p}\times\underbrace{\frac{R_i-1}{N_i}}_{\to p}\times\cdots\times\underbrace{\frac{R_i-k+1}{N_i}}_{\to p}\right]\cdot \quad \leftarrow N_i^k$$

$$N_i^{n-k} \quad \left[\underbrace{\frac{N_i-R_i}{N_i}}_{\to 1-p}\times\underbrace{\frac{(N_i-R_i)-1}{N_i}}_{\to 1-p}\times\cdots\times\underbrace{\frac{(N_i-R_i)-(n-k)+1}{N_i}}_{\to 1-p}\right]\cdot$$

$$N_i^{n} \quad \left[\underbrace{\frac{N_i}{N_i}}_{\to 1}\times\underbrace{\frac{N_i}{N_i-1}}_{\to 1}\times\cdots\times\underbrace{\frac{N_i}{N_i-n+1}}_{\to 1}\right]$$

$$\to \binom{n}{k}p^k(1-p)^{n-k}$$

- Summary for $X \sim$ Hypergeometric(n, N, R)
  - Range: $\mathcal{X} = \{0, 1, 2, \ldots, n\}$ → $\max(0, n+R-N) \le X \le \min(R, n)$
  - Pmf: $f_X(x) = \binom{R}{x}\binom{N-R}{n-x}/\binom{N}{n}$, for $x \in \mathcal{X}$
  - Parameters: n, N, $R \in \{1, 2, 3, \ldots\}$ and $n \le N$, $R \le N$
  - Mean: $E(X)=nR/N$
  - Variance: $Var(X)=nR(N-R)(N-n)/(N^2(N-1))$

❖ **Reading**: textbook, Sec 4.6, 4.7, 4.8.1~4.8.3