# Statistikk

Jo Eidsvik

Matematiske fag, NTNU

# Diskrete fordelinger

Stokastisk variabel $X$. Punktsannsynlighet $f(x)$.

$$\sum_{x \in S} f(x) = 1, \qquad f(x) \geq 0.$$

# Geometrisk fordeling

Antakelser:

- uavhengige forsøk.
- konstant sannsynlighet for suksess $p$ i alle forsøk.

$X$ = antall forsøk til første suksess.

$$f(x) = (1-p)^{x-1}p, \quad x = 1, \ldots$$

# Forventning: Geometrisk fordeling

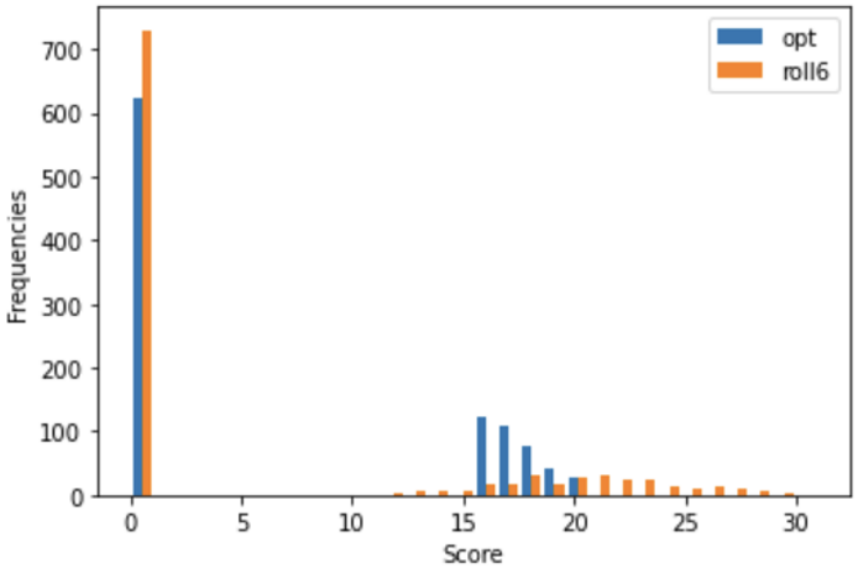
La $q = 1 - p$. Da er $E(X) = \sum_{x=1}^{\infty} xq^{x-1}p$.

$$E(X) = p\sum_{x=1}^{\infty} xq^{x-1} = p\sum_{x=1}^{\infty} \frac{d}{dq}q^x = p\frac{d}{dq}\sum_{x=1}^{\infty} q^x$$

$$= p\frac{d}{dq}\frac{q}{1-q} = p\frac{1}{(1-q)^2} = 1/p$$

# Terningkast, score til 0 om 6 trilles

I et spill kastes terningen gjentatte ganger, og man forsøker å oppnå høy score. La score etter kast $k$ være $Z_k$. Før kast $k+1$ velger man selv om det skal kastes et kast til, eller om spillet skal avsluttes og da beholder man sin nåværende score $Z_k$. Man starter spillet med score $Z_0 = 0$ ved $k = 0$. Dersom utfallet av kast $k+1$ er $x_{k+1} \in \{1, \ldots, 5\}$, så adderes utfallet til scoren $(Z_{k+1} = Z_k + X_{k+1})$. Dersom utfallet blir $x_{k+1} = 6$, så settes scoren $Z_{k+1} = 0$, og spillet avsluttes.




opt = Kast hvis forventet ny score er større
roll6= Kast max 6 kast

```
Strategy stop 15:
6489
Strategy 6 rolls:
5591
```

# Negativ binomisk fordeling

Antakelser:

- uavhengige forsøk.
- konstant sannsynlighet for suksess $p$ i alle forsøk.

*Negativ binomisk er en utvidelse av geometrisk: Antall forsøk til $k$-te suksess.*

# Poisson fordeling

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, \ldots$$

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

# Poisson fordeling for tid $(0, t)$

$$f(x) = e^{-\lambda t}\frac{(\lambda t)^x}{x!}, \quad x = 0, 1, \ldots$$

$$E(X) = \lambda t$$

$$Var(X) = \lambda t$$

# Eksempel: Krav til forsikringselskap




- Forsikringselskap må dokumentere at de har nok penger til å betale ut krav.
- Inn: Premie * antall kunder. Ut: sum av inkommende krav
- Antar samme beløp i hvert krav, så ser bare på antall krav.
- $X =$ antall krav i tid $t$. Poissonfordelt med parameter $\lambda t$, $\lambda = 10$, $t = 1$ for 1 år.

Hva er sannsynligheten for at det kommer flere enn 20 krav på 1 år?

# Eksempel: Krav til forsikringselskap

$\lambda = 10$

$$P(X > 20) = 1 - P(X \leq 20) = 1 - F(20) = 1 - \sum_{x=0}^{20} e^{-10} \frac{10^x}{x!}$$

**Table A.2** (continued) Poisson Probability Sums $\sum_{x=0}^{r} p(x;\mu)$

| | $\mu$ | | | | | |
|---|---|---|---|---|---|---|
| $r$ | 10.0 | 11.0 | 12.0 | 13.0 | 14.0 | 15.0 |
| 0 | 0.0000 | 0.0000 | 0.0000 | | | |
| 1 | 0.0005 | 0.0002 | 0.0001 | 0.0000 | 0.0000 | |
| 2 | 0.0028 | 0.0012 | 0.0005 | 0.0002 | 0.0001 | 0.0000 |
| 3 | 0.0103 | 0.0049 | 0.0023 | 0.0011 | 0.0005 | 0.0002 |
| 4 | 0.0293 | 0.0151 | 0.0076 | 0.0037 | 0.0018 | 0.0009 |
| 5 | 0.0671 | 0.0375 | 0.0203 | 0.0107 | 0.0055 | 0.0028 |
| 6 | 0.1301 | 0.0786 | 0.0458 | 0.0259 | 0.0142 | 0.0076 |
| 7 | 0.2202 | 0.1432 | 0.0895 | 0.0540 | 0.0316 | 0.0180 |
| 8 | 0.3328 | 0.2320 | 0.1550 | 0.0998 | 0.0621 | 0.0374 |
| 9 | 0.4579 | 0.3405 | 0.2424 | 0.1658 | 0.1094 | 0.0699 |
| 10 | 0.5830 | 0.4599 | 0.3472 | 0.2517 | 0.1757 | 0.1185 |
| 11 | 0.6968 | 0.5793 | 0.4616 | 0.3532 | 0.2600 | 0.1848 |
| 12 | 0.7916 | 0.6887 | 0.5760 | 0.4631 | 0.3585 | 0.2676 |
| 13 | 0.8645 | 0.7813 | 0.6815 | 0.5730 | 0.4644 | 0.3632 |
| 14 | 0.9165 | 0.8540 | 0.7720 | 0.6751 | 0.5704 | 0.4657 |
| 15 | 0.9513 | 0.9074 | 0.8444 | 0.7636 | 0.6694 | 0.5681 |
| 16 | 0.9730 | 0.9441 | 0.8987 | 0.8355 | 0.7559 | 0.6641 |
| 17 | 0.9857 | 0.9678 | 0.9370 | 0.8905 | 0.8272 | 0.7489 |
| 18 | 0.9928 | 0.9823 | 0.9626 | 0.9302 | 0.8826 | 0.8195 |
| 19 | 0.9965 | 0.9907 | 0.9787 | 0.9573 | 0.9235 | 0.8752 |
| 20 | 0.9984 | 0.9953 | 0.9884 | 0.9750 | 0.9521 | 0.9170 |
| 21 | 0.9993 | 0.9977 | 0.9939 | 0.9859 | 0.9712 | 0.9469 |
| 22 | 0.9997 | 0.9990 | 0.9970 | 0.9924 | 0.9833 | 0.9673 |
| 23 | 0.9999 | 0.9995 | 0.9985 | 0.9960 | 0.9907 | 0.9805 |
| 24 | 1.0000 | 0.9998 | 0.9993 | 0.9980 | 0.9950 | 0.9888 |
| 25 | | 0.9999 | 0.9997 | 0.9990 | 0.9974 | 0.9938 |
| 26 | | 1.0000 | 0.9999 | 0.9995 | 0.9987 | 0.9967 |

# Eksempel: Krav til forsikringselskap

$$P(X > 20) = 1 - P(X \leq 20) = 1 - F(20) = 1 - 0.9984 = 0.0016,$$

Dette er en liten sannsynlighet.

Slike beregninger brukes til å godkjenne forsikringsaktivitet, regne ut premier, etc.

# Eksempel: Leukemi i Niles, Illinois, USA

TABLE I

INCIDENCE OF LEUKEMIA AMONG WHITE CHILDREN UNDER AGE FIFTEEN IN NILES, ILLINOIS, AND VICINITY, 1956–1960

| Area | Population Under Age 15 (1960 Census) | | Cases of Leukemia in Children Under Age 15* (1956–1960) | | | Average Annual Rate per 100,000 | |
|---|---|---|---|---|---|---|---|
| | | | Recorded Deaths | Physician Reports† | Total | | |
| Niles | | 7,076 | 2‡ | 6 | 8‡ | | 21.3 |
| Neighboring towns | | 44,075 | 10 | 3 | 13 | | 5.6 |
| Des Plaines | 11,952 | | 3 | ... | 3 | 4.7 | |
| Lincolnwood | 3,654 | | 1 | ... | 1 | 5.2 | |
| Morton Grove | 7,995 | | 2 | 2 | 4 | 9.4 | |
| Skokie | 20,474 | | 4 | 1 | 5 | 4.6 | |
| Cook County, Illinois, exclusive of Niles | | 1,152,695 | 241 | 45 | 286 | | 4.6 |

Over 5 år:

- Niles: 8 tilfeller av totalt 7076 barn under 15 år.
- Illinois: 286 tilfeller av totalt 1.152 million barn under 15 år.

Er Niles 'annerledes' - forurenset vann eller annet?

# Eksempel: Leukemi i Niles, Illinois, USA

I befolkning: $p = 286/1.152mill = 0.000248$
Forventet verdi (intensitet) i Niles: $\lambda = np = 7076 \cdot 0.000248 = 1.75$

$$P(X \geq 8) = 1 - P(X \leq 7) = 1 - F(7) = 1 - \sum_{x=0}^{7} e^{-1.75} \frac{1.75^x}{x!} = 0.0005 = 1/2000$$

Dette er en liten sannsynlighet. Må undersøkes mer.

# Link fordelinger

- Binomisk med $n$ veldig stor, og $p$ veldig liten : Poissonfordelingen $\lambda = np$ er god approksimasjon

Tidligere: Hypergeometrisk: $n$ trekninger liten mot antall muligheter $N$ : Binomisk fordeling er god approksimasjon
(trekning med eller uten tilbakelegging er nokså likt)

# Dear Abby, 266 days

Dear Abby: You wrote in your column that a woman is pregnant for 266 days. Who said so? I carried my baby for ten months and five days, and there is no doubt about it because I know the exact date my baby was conceived. My husband is in the Navy and it couldn't have possibly been conceived any other time because I saw him only once for an hour, and I didn't see him again until the day before the baby was born.

I don't drink or run around, and there is no way this baby isn't his, so please print a retraction about the 266-day carrying time because otherwise I am in a lot of trouble.

San Diego Reader

# Dear Abby, 266 days

Lengde på svangerskap:

- $\mu = 266$
- $\sigma = 16$
- Normalfordelt

Er 310 dager veldig lenge?