Funktioner av stokastiska variabler

Funktioner av stokastiska variabler är nya stokastiska variabler. Låt X vara en stokastisk variabel och $g(x)$ en reellvärd funktion. Då är $Z = g(X)$ en stokastisk variabel.

Exempel: X och Y oberoende diskreta stokastiska variabler. Låt $Z = X + Y$. Oberoendet ger att $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ och

$$\mathsf{P}(X + Y = n) = \sum_k \mathsf{P}(X = k)\,\mathsf{P}(Y = n - k)$$

I det kontinuerliga fallet får man motsvarande

$$f_{X+Y}(z) = \int f_X(t)\, f_Y(z - t)\, dt$$

Formeln kallas *faltningsformeln*.

Exempel: Låt $X_1, \ldots, X_n$ vara oberoende, likafördelade stokastiska variabler med fördelningsfunktion $F_X(t)$. Med $Z = \max(X_1, \ldots, X_n)$ så

$$\begin{aligned} F_Z(t) &= \mathsf{P}(Z \le t) = \mathsf{P}(\max(X_1, \ldots, X_n) \le t) = \mathsf{P}(X_1 \le t, \ldots, X_n \le t) \\ &= \mathsf{P}(X_1 \le t) \cdots \mathsf{P}(X_n \le t) = (F_X(t))^n \end{aligned}$$

Väntevärden och andra parametrar

Beskrivande parametrar för en stokastisk variabel:

- Genomsnittligt värde, väntevärde $\mathsf{E}[X]$
- Spridningsmått, varians $\mathrm{V}(X)$ och standardavvikelse $\mathrm{D}(X)$

Definition: *Väntevärdet* för en stokastisk variabel, $\mathsf{E}[X]$, definieras som

$$\mathsf{E}[X] = \begin{cases} \displaystyle\sum_{k \in S_X} k \cdot \mathsf{P}(X = k) & \text{om } X \text{ är diskret} \\ \displaystyle\int_{-\infty}^{\infty} x \cdot f_X(x)\, dx & \text{om } X \text{ är kontinuerlig} \end{cases}$$

Väntevärdet är ett teoretiskt genomsnittligt värde.

Exempel: X beskriver antalet tärningskast tills första 5:a. X är ffg(p), $p = 1/6$.

$$p_X(k) = \mathsf{P}(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \ldots$$

Dår är

$$
\begin{aligned}
\mathsf{E}[X] &= \sum_k k\mathsf{P}(X=k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p + \sum_{k=2}^{\infty} k(1-p)^{k-1}p \\
&= p + \sum_{j=1}^{\infty} (j+1)(1-p)^j p \\
&= p + (1-p)\sum_{j=1}^{\infty} j \underbrace{(1-p)^{j-1}p}_{=p_X(j)} + (1-p)\sum_{j=1}^{\infty} \underbrace{(1-p)^{j-1}p}_{=p_X(j)} \\
&= p + (1-p)\underbrace{\sum_{j=1}^{\infty} j p_X(j)}_{=\mathsf{E}[X]} + (1-p)\underbrace{\sum_{j=1}^{\infty} p_X(j)}_{=1} \\
&= p + (1-p)\mathsf{E}[X] + (1-p) = (1-p)\mathsf{E}[X] + 1
\end{aligned}
$$

vilket ger

$$
\mathsf{E}[X] = \frac{1}{p}.
$$

Exempel: X är resultatet av ett tärningskast. $\mathsf{P}(X=i) = 1/6$ för $i = 1, 2, \ldots, 6$.

$$
\mathsf{E}[X] = \sum_{k \in S_X} k\mathsf{P}(X=k) = \sum_{k=1}^{6} k\frac{1}{6} = \frac{1}{6} + \frac{2}{6} + \cdots \frac{6}{6} = 3.5.
$$

Exempel: X är exponentialfördelad, dvs $f_X(x) = \frac{1}{m}\mathrm{e}^{-x/m}$ för $x \geq 0$.

$$
\mathsf{E}[X] = \int_0^{\infty} x \cdot \frac{1}{m}\mathrm{e}^{-x/m}\,dx = \cdots = m.
$$

Sats (Den aningslöse statistikerns sats). För en reellvärd funktion $g(x)$ och en stokastisk variabel X är

$$
\mathsf{E}[g(X)] = \begin{cases} \displaystyle\sum_{k \in S_X} g(k) \cdot \mathsf{P}(X=k) & \text{om } X \text{ är diskret} \\ \displaystyle\int_{-\infty}^{\infty} g(x) \cdot f_X(x)\,dx & \text{om } X \text{ är kontinuerlig} \end{cases}
$$

förutsatt att summan/integralen är absolutkonvergent.

Spridningsmått?

Definition: *Variansen* för en stokastisk variabel definieras och betecknas

$$
\mathrm{V}(X) = \mathsf{E}\left[(X - \mathsf{E}[X])^2\right].
$$

Definition: *Standardavvikelsen* för en stokastisk variabel definieras och betecknas

$$
\mathrm{D}(X) = \sqrt{\mathrm{V}(X)}.
$$

Denna har rätt enhet.

Exempel: Tärningskast

$$\mathrm{V}(X) = \mathsf{E}\left[(X-3.5)^2\right] = \sum_{k=1}^{6}(k-3.5)^2\frac{1}{6} = \frac{35}{12} \approx 2.917$$

(enhet: ”kvadratprickar”)

Exempel: Tärningskast

$$\mathrm{D}(X) = \sqrt{\mathrm{V}(X)} = \sqrt{\frac{35}{12}} \approx 1.71$$

(enhet: ”prickar”)

En liten insyn i vad standardavvikelsen betyder kan fås genom följande.

Sats (Tjebychevs olikhet). För $k > 0$ är

$$\mathsf{P}(|X - \mathsf{E}[X]| > k\mathrm{D}(X)) < \frac{1}{k^2}.$$

Bevis: (kontinuerliga fallet)

$$\int\limits_{\frac{|x-m|}{k\sigma}>1} f_X(x)\,dx < \int\limits_{\frac{|x-m|}{k\sigma}>1} \left(\frac{x-m}{k\sigma}\right)^2 f_X(x)\,dx \leq \int\limits_{-\infty}^{\infty} \frac{(x-m)^2}{k^2\sigma^2} f_X(x)\,dx = \frac{1}{k^2\sigma^2}\mathrm{V}(X) = \frac{1}{k^2}.$$

Sats. Låt X vara en stokastisk variabel och a och b två konstanter. Då är

$$\mathsf{E}[aX+b] = a\mathsf{E}[X] + b \qquad \mathrm{V}(aX+b) = \mathrm{V}(aX) = a^2\mathrm{V}(X).$$

Här bevisas det kontinuerliga fallet

$$\begin{aligned}
\mathsf{E}[aX+b] &= \int_{-\infty}^{\infty} g(x)f_X(x)\,dx = \int_{-\infty}^{\infty}(ax+b)f_X(x)\,dx \\
&= a\int_{-\infty}^{\infty} xf_X(x)\,dx + b\int_{-\infty}^{\infty} f_X(x)\,dx = a\mathsf{E}[X] + b
\end{aligned}$$

Diskreta fallet är analogt. Vidare, eftersom $\mathsf{E}[aX+b] = a\mathsf{E}[X] + b$ så är

$$\begin{aligned}
\mathrm{V}(aX+b) &= \mathsf{E}\left[((aX+b) - \mathsf{E}[aX+b])^2\right] = \mathsf{E}\left[(aX+b-a\mathsf{E}[X]-b)^2\right] \\
&= \mathsf{E}\left[a^2(X-\mathsf{E}[X])^2\right] = a^2\mathsf{E}\left[(X-\mathsf{E}[X])^2\right] = a^2\mathrm{V}(X)
\end{aligned}$$

Nästa tillfälle skall följande beräkningsaspekt för varianser bevisas:

$$\mathrm{V}(X) = \mathsf{E}\left[X^2\right] - (\mathsf{E}[X])^2.$$