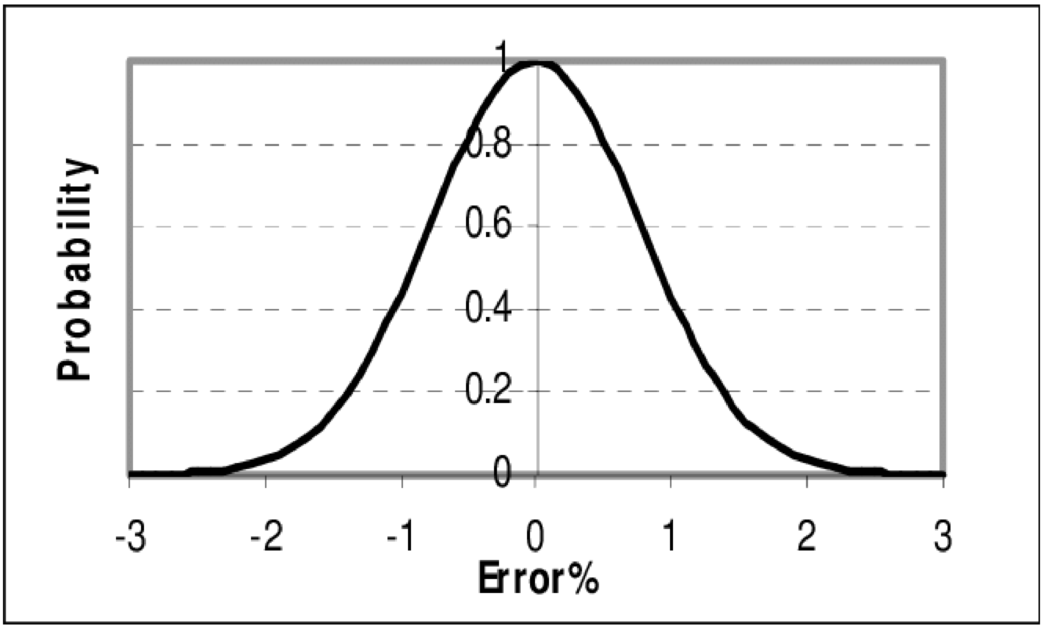

Figure 0.1: Normal distribution of theerror term

تفسير ليه بنستخدم دالة الخطأ المربعة كدالة تكلفة في Linear Regression

١. الفرضية الأساسية

لما بنعمل نموذج Linear Regression، إحنا بنفترض إن العلاقة بين المتغيرات x والمتغير اللي عايزين نتوقعه y ممكن تتكتب كالتالي:

$$y = \theta^T x + \epsilon$$

شرح المعادلة:

- y: القيمة اللي عايزين نتوقعها (مثلاً سعر شقة).
- $\theta^T x$: ده التوقع بتاعنا باستخدام النموذج. القيم θ هي معاملات النموذج اللي عايزين نتعلمها.
- ϵ: الخطأ، الفرق بين القيمة الحقيقية y والقيمة المتوقعة $\theta^T x$.

إحنا بنفترض إن الخطأ ϵ بيتوزع توزيع طبيعي (Gaussian Distribution) بمتوسط 0 وتباين σ^2:

$$\epsilon \sim N(0, \sigma^2)$$

ده معناه ايه؟ واية اللي احنا بنفرضوا لما بنفترض التوزيع بتاعنا بالشكل ده؟

- الأخطاء الصغيرة شائعة أكتر من الأخطاء الكبيرة.
- الأخطاء بتكون موزعة حوالين القيمة صفر بمعني ان المتوسط بتاع الخطاء هو صفر.
- ـ التباين σ^2 بيحدد مدى انتشار الأخطاء: لو صغير، الأخطاء قريبة من الصفر. لو كبير، الأخطاء متوزعة أكتر.

---

٢. الاحتمال الشرطي

الصيغة الرياضية:

$$p(y|x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\theta^T x)^2}{2\sigma^2}\right)$$

إيه معنى الاحتمال الشرطي؟

الاحتمال الشرطي بيقولنا: إيه احتمالية إن القيمة y الحقيقية تظهر، بشرط إننا عارفين القيم x اللي بنستخدمها كمدخلات والقيم θ اللي بتحدد النموذج

- $\theta^T x$: **هو التوقع اللي بنحسبه باستخدام النموذج.**
- $(y-\theta^T x)$: **هو الفرق بين القيمة الحقيقية y والتوقع، أو بمعنى تاني الخطأ ϵ.**
- **التوزيع الطبيعي (Gaussian): بيفترض إن الخطأ ϵ بيكون صغير في معظم الأحيان وقريب من ٠.**

تفسير رياضي:

- **الاحتمال الشرطي بيوضح لنا مدى قرب y من التوقع $\theta^T x$.**
- **لو y قريب من $\theta^T x$، القيمة $(y-\theta^T x)^2$ هتكون صغيرة، وبالتالي $p(y|x;\theta)$ هيكون كبير.**
- **لو y بعيد عن $\theta^T x$، القيمة $(y-\theta^T x)^2$ هتكون كبيرة، وده هيخلي $p(y|x;\theta)$ صغير.**

---

٣. تعظيم دالة الاحتمال (Likelihood)

لما يكون عندنا أكتر من نقطة بيانات n، دالة الاحتمال الإجمالية هي:

$$L(\theta) = \prod_{i=1}^{n} p(y^{(i)}|x^{(i)};\theta)$$

إيه اللي دالة الاحتمال بتعمله؟

- دالة الاحتمال بتقيس احتمالية إن كل القيم الحقيقية $y^{(i)}$ تظهر بالنظر إلى القيم $x^{(i)}$ والنموذج بتاعنا اللي بنحدده من خلال θ.
- الهدف الأساسي هو تعظيم $L(\theta)$: يعني نختار القيم θ اللي تخلي النموذج يمثل البيانات بأكبر احتمالية ممكنة.

ليه بنهتم بدالة الاحتمال؟

- تعظيم $L(\theta)$ معناه إننا بنحسن توقعات النموذج بحيث تكون أقرب ما يمكن للبيانات الحقيقية.
- دالة الاحتمال بتربط بين الفرضية الإحصائية للنموذج (التوزيع الطبيعي) والبيانات اللي عندنا.

ليه بناخد اللوغاريتم (Log-Likelihood)؟

بدل ما نشتغل مع دالة الاحتمال مباشرة، بناخد اللوغاريتم الطبيعي عشان الحساب يكون أسهل:

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log p(y^{(i)}|x^{(i)};\theta)$$

١. **ليه؟**

اللوغاريتم بيحول الضرب إلى جمع، وده بيخلي الاشتقاقات والحسابات أسهل.

٢. **إيه الهدف؟**

الهدف لسه هو نفسه: نختار القيم θ اللي تعظم اللوغاريتم.
إيه اللي بنوصله بعد كده؟

لما نحط $p(y|x;\theta)$ في اللوغاريتم ونفصل المعادلة، بنلاقي:

$$\ell(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y^{(i)} - \theta^T x^{(i)})^2$$

الجزء المهم هنا هو الترم التاني، لأنه بيعبر عن الخطأ بين القيم الحقيقية والتوقعات.

النتيجة النهائية: عشان نعظم اللوغاريتم، بنقلل الترم ده:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} (y^{(i)} - \theta^T x^{(i)})^2$$

ده هو السبب الأساسي لاختيار دالة الخطأ المربعة كدالة تكلفة في Linear Regression.